



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification:</b> <b>C12N 15/10, A01H 5/00,</b> <b>C12N 5/10, C12N 15/54,</b> <b>C12Q 1/48, C12Q 1/68</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 00/28018</b> <b>(43) International Publication Date:</b> 18 May 2000 (18.05.2000)
<b>(21) International Application Number:</b> PCT/US99/26797 <b>(22) International Filing Date:</b> 09 November 1999 (09.11.1999) <b>(30) Priority Data:</b> 60/107,782 10 November 1998 (10.11.1998) US <b>(60) Parent Application or Grant</b> MAXYGEN, INC. [/]; (). STEMMER, Willem, P., C. [/]; (). SUBRAMANIAN, Venkiteswaran [/]; (). STEMMER, Willem, P., C. [/]; (). SUBRAMANIAN, Venkiteswaran [/]; (). QUINE, Jonathan, Alan ; ().		<b>Published</b>
<b>(54) Title: MODIFIED ADP-GLUCOSE PYROPHOSPHORYLASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES</b> <b>(54) Titre: ADP-GLUCOSE PYROPHOSPHORYLASE MODIFIEE POUR L'AMELIORATION ET L'OPTIMISATION DE PHENOTYPES VEGETAUX</b>  <b>(57) Abstract</b> The invention provides methods and compositions relating to sequence-shuffled variants of ADP-glucose pyrophosphorylase.  <b>(57) Abrégé</b> Cette invention porte sur des procédés et des compositions relatifs à la manipulation de séquences de variantes de l'ADP-glucose pyrophosphorylase.		

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>7</sup> : C12N 15/10, C12Q 1/48, 1/68, C12N 5/10, A01H 5/00, C12N 15/54		A1	(11) International Publication Number: <b>WO 00/28018</b> (43) International Publication Date: 18 May 2000 (18.05.00)
(21) International Application Number: PCT/US99/26797 (22) International Filing Date: 9 November 1999 (09.11.99) (30) Priority Data: 60/107,782 10 November 1998 (10.11.98) US (71) Applicant (for all designated States except US): MAXYGEN, INC. [US/US]; 515 Galveston Drive, Redwood City, CA 94063 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): STEMMER, Willem, P., C. [NL/US]; 108 Kathy Court, Los Gatos, CA 95032 (US). SUBRAMANIAN, Venkiteswaran [US/US]; 3980 Corte Mar de Hierba, San Diego, CA 92130 (US). (74) Agents: QUINE, Jonathan, Alan; The Law Offices of Jonathan Alan Quine, P.O. Box 458, Alameda, CA 94501 (US) et al.			(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: MODIFIED ADP-GLUCOSE PYROPHOSPHORYLASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES			
(57) Abstract  The invention provides methods and compositions relating to sequence-shuffled variants of ADP-glucose pyrophosphorylase.			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

Description

5

10

15

20

25

30

35

40

45

50

55

5

10

5

Modified ADP-Glucose Pyrophosphorylase  
for Improvement and Optimization of Plant Phenotypes

CROSS REFERENCE TO RELATED APPLICATIONS

15

10

This application is a non-provisional filing of and claims priority to provisional patent application "MODIFIED ADP-GLUCOSE PYROPHOSPHORYLASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES" by Willem P.C. Stemmer and Venkiteswaran Subramanian, USSN 60/107,782, filed November 10, 1998.

20

FIELD OF THE INVENTION

25

15

The invention relates to methods and compositions for generating, modifying, adapting, and optimizing polynucleotide sequences that encode proteins having ADPGPP enzyme activities which are useful for introduction into plant species, and other hosts, and related aspects.

BACKGROUND

30

Genetic Engineering of Plants

20

35

40

45

25

30

Genetic engineering of agricultural organisms dates back thousands of years to the dawn of agriculture. The hand of man has selected the agricultural organisms having the phenotypic traits that were deemed desirable, which desired phenotypic traits have often been taste, high yield, caloric value, ease of propagation, resistance to pests and disease, and appearance. Classical breeding methods to select for germplasm encoding desirable agricultural traits had been a standard practice of the world's farmers long before Gregor Mendel and others identified the basic rules of segregation and selection. For the most part, the fundamental process underlying the generation and selection of desired traits was the natural mutation frequency and recombination rates of the organisms, which are quite slow compared to the human lifespan and make it difficult to use conventional methods of breeding to rapidly obtain or optimize desired traits in an organism.

50

55

5 The very recent advent of non-classical, or "recombinant" genetic  
engineering techniques has provided a new means to expedite the generation of  
10 agricultural organisms having desired traits that provide an economic, ecological,  
nutritional, or aesthetic benefit. To date, most recombinant approaches have involved  
15 transferring a novel or modified gene into the germline of an organism to effect its  
expression or to inhibit the expression of the endogenous homologue gene in the  
organism's native genome. However, the currently used recombinant techniques are  
generally unsuited for substantially increasing the rate at which a novel or improved  
20 phenotypic trait can be evolved. Essentially all recombinant genes in use today for  
agriculture are obtained from the germplasm of existing plant and microbial  
specimens, which have naturally evolved coordinately with constraints related to  
other aspects of the organism's evolution and typically are not optimized for the  
desired phenotype(s). The sequence diversity available is limited by the natural  
25 genetic variability within the existing specimen gene pool, although crude mutagenic  
approaches have been used to add to the natural variability in the gene pool.

Unfortunately, the induction of mutations to generate diversity often  
requires chemical mutagenesis, radiation mutagenesis, tissue culture techniques, or  
mutagenic genetic stocks. These methods provide means for increasing genetic  
30 variability in the desired genes, but frequently produce deleterious mutations in many  
other genes. These other traits may be removed, in some instances, by further genetic  
manipulation (e.g., backcrossing), but such work is generally both expensive and time  
consuming. For example, in the flower business, the properties of stem strength and  
35 length, disease resistance and maintaining quality are important, but often initially  
compromised in the mutagenesis process.

#### 25 ADP-Glucose Pyrophosphorylase

40 The biosynthesis of starches in higher plants occurs in three steps, the  
first of which involves synthesis of ADP glucose from ATP and  $\alpha$ -glucose-1-  
phosphate, and which is catalyzed by ADP-glucose pyrophosphorylase ("ADPGPP";  
EC 2.7.7.27) The second step of starch biosynthesis is transfer of a glucosyl moiety  
45 of ADP-glucose to a maltodextrin or starch to give rise to a new  $\alpha$ -1,4-glucosyl  
linkage; the reaction is catalyzed by a starch synthase, of which there are several  
forms present either as soluble enzymes or bound to starch particles as particulate

5 enzymes. The third reaction is catalyzed by branching enzyme and is responsible for synthesis of  $\alpha$ -1,6-glucosyl linkages.

10 Starch synthesis in plants is tightly regulated and is tied to photosynthetic carbon fixation. Principal control of starch synthesis in plants, algae, and bacteria is at the level of ADPGPP. It has been shown that reduced ADPGPP activity in *Arabidopsis* leaves and potato tubers results in a reduced rate of starch synthesis. The ADPGPP enzyme in plants exists primarily as a tetramer,  $S_2L_2$ , composed of two different subunits of approximately 50-60 kDa each. The molecular weight of the small (S) subunit is approximately 50-55 kDa, and the S subunit is the catalytic protein having the enzymatic active site (e.g., reaction center). The molecular weight of the large (L) subunit is approximately 55-60 kDa, and the L subunit is the regulatory subunit protein. The plant enzyme is strongly inhibited by 3-phosphoglycerate (PGA), a product of carbon dioxide fixation; in the absence of PGA, the enzyme exhibits only about 3% of its activity. Plant ADPGPP is also strongly inhibited by inorganic phosphate (Pi). In contrast, bacterial and algal ADPGPP exist as homotetramers of 50kDa. The Algal enzyme, like its plant counterpart, is activated by PGA and inhibited by Pi, whereas the bacterial enzyme is activated by fructose-1,6-bisphosphate (FBP) and inhibited by AMP and Pi.

15 In the last 10 years, the demand for starch has dramatically increased both for food and industrial uses, primarily as a result of increased demand for high fructose corn syrups and biofuel. Hence, mobilizing a greater proportion of the photosynthetic assimilates of major crops into the seeds and other sinks in the form of starch can be expected to have a major impact on agriculture in the form of increased yield of harvestable parts. Deregulating starch biosynthesis by deregulating ADPGPP (e.g., decoupling from the need for positive activation and/or negative inhibition of catalytic activity) in order to increase both the rate of accumulation and the amount of starch in sinks such as tubers (e.g., potato) and seeds (e.g., maize, wheat, rice). A mutant form of *E. coli* ADPGPP gene (Glg C16) has been introduced into potato and exhibits a significant activity in the absence of its normal activator, FGP, and is much less sensitive to feedback inhibition by AMP and Pi. Transgenic potato plants expressing this gene under the control of a tuber-specific promoter showed 25-60% more starch in tubers as compared to control non-transgenic plants.

5 As noted, the advent of recombinant DNA technology has provided  
agriculturists with additional means of modifying plant genomes. While certainly  
practical in some areas, to date genetic engineering methods have had limited success  
in transferring or modifying important biosynthetic or other pathways, including the  
10 5 ADPGPP enzyme in photosynthetic organisms and bacteria. The creation of plants  
and other photosynthetic organisms having improved ADPGPP biosynthetic pathways  
can provide increased yields of certain types of starchy foodstuffs, enhanced biomass  
energy sources, and may alter the types and amounts of nutrients present in certain  
15 foodstuffs, among other desirable phenotypes.

10 Thus, there exists a need for improved methods for producing plants  
and agricultural photosynthetic microbes with an improved ADPGPP enzyme. In  
particular, these methods should provide general means for producing novel ADPGPP  
enzymes, including increasing the diversity of the ADPGPP gene pool and the rate at  
which genetic sequences encoding one or more ADPGPP subunits having desired  
25 15 properties are evolved. It is particularly desirable to have methods which are suitable  
for rapid evolution of genetic sequences to function in one or more plant species and  
confer an improved ADPGPP phenotype (e.g., reduced sensitivity to inhibitors (e.g.,  
Pi, AMP), reduced dependence on activators (e.g., PGA, FBP), improved catalytic  
efficiency via increasing Vmax and/or increasing the apparent affinity of substrates  
30 20 for the enzyme, and/or relieving a requirement for allosteric activation or inhibition  
by allosteric repression, as well as plants which express the novel ADPGPP genetic  
sequence(s).

35 The present invention meets these and other needs and provides such  
improvements and opportunities.

25 The references discussed herein are provided solely for their disclosure  
prior to the filing date of the present application. Nothing herein is to be construed as  
an admission that the inventors are not entitled to antedate such disclosure by virtue  
of prior invention. All publications cited are incorporated herein by reference,  
whether specifically noted as such or not.

#### 45 30 SUMMARY OF THE INVENTION

In a broad general aspect, the present invention provides a method for  
the rapid evolution of one or more polynucleotide sequences encoding a ADPGPP



5 enzyme, or one or more subunits thereof, that, when transferred into an appropriate  
plant cell, or photosynthetic microbial host and expressed therein, confers an  
enhanced metabolic phenotype to the host to increase starch formation ratio and/or  
rate, or to increase the accumulation or depletion of certain starches. In general,  
10 polynucleotide sequence shuffling and phenotype selection, such as detection of a  
parameter of ADPGPP enzyme activity, is employed recursively to generate  
polynucleotide sequences which encode novel proteins having desirable ADPGPP  
enzymatic catalytic function(s), regulatory function(s), and related enzymatic and  
15 physicochemical properties. Although the method is believed broadly applicable to  
evolving biosynthetic enzymes having desired properties, the invention is described  
principally with reference to the metabolic enzyme activities of plants and/or  
20 photosynthetic microbes and/or bacteria, defined as ADPGPP, or an isozyme thereof,  
including both catalytic subunit (small subunit, S; gene designation, S) and allosteric  
regulatory subunit (large subunit, L; gene designation, L), respectively, as appropriate  
25 for plant and algal ( $S_2L_2$ ), as well as bacterial ( $S_4$ ).

In one aspect, the invention provides methods of producing a  
recombinant cell having an elevated starch production activity. In the methods, one  
or more first ADGPP enzyme coding nucleic acid, or a homologue thereof, is  
30 recombined with one or more homologous first nucleic acid to produce a library of  
recombinant first enzyme nucleic acid homologues. This step can be repeated as  
desired to produce a more diverse library of recombinant first enzyme nucleic acid  
homologues. The libraries are selected for an activity which aids in Starch  
35 production, such as an increased or decreased catalytic rate, an altered substrate  
specificity, an increased ability of a cell expressing one or more members of the  
library to produce starch when the one or more library members is expressed in the  
40 cell, etc., thereby producing a selected library of recombinant first enzyme nucleic  
acid homologues. These steps are recursively repeated until one or more members of  
the selected library produces an elevated starch production level in a target  
recombinant cell when the one or more selected library member is expressed in the  
45 target cell, as compared to a starch fixation activity of the target cell when the one or  
more selected library member is not expressed in the target cell.

Other features and advantages of the invention will be apparent from the following description of the drawings, preferred embodiments of the invention, the examples, and the claims.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Desensitization of ADPGPP to activator and inhibitor. Panel A shows a diagrammatic representation of ADPGPP activity as a function of activator concentration for a parental wild-type ADPGPP (solid line), a shufflant which is partially desensitized (dotted line), and a shufflant which is fully desensitized (dashed line) to activator. Panel B shows a diagrammatic representation of ADPGPP activity as a function of inhibitor concentration for a parental wild-type ADPGPP (solid line), a shufflant which is partially desensitized (dotted line), and a shufflant which is fully desensitized (dashed line) to inhibitor.

Figure 2. Optimization by shuffling of ADPGPP for substrate usage and resistance to inhibition. Panel A shows a diagrammatic representation of ADPGPP activity as a function of substrate concentration for a parental wild-type ADPGPP (solid line), and a shufflant which is optimized for substrate usage (dashed line);  $K_m$  for the wildtype  $K_m(wt)$  and optimized enzyme  $K_m(opt)$ , and  $V_{max}$  for the wildtype  $V_{max}(wt)$  and optimized  $V_{max}(opt)$  are shown. Panel B shows a diagrammatic representation of ADPGPP activity as a function of inhibitor concentration for a parental wild-type ADPGPP (solid line), and a shufflant which is optimized for substrate usage (dashed line);  $K_m$  for the wildtype  $K_m(wt)$  and optimized enzyme  $K_m(opt)$ , and  $V_{max}$  for the wildtype  $V_{max}(wt)$  and optimized  $V_{max}(opt)$  are shown.

#### DETAILED DESCRIPTION

##### Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. For purposes of the present invention, the following terms are defined below.

5                   The term "shuffling" is used herein to indicate recombination between  
substantially homologous but non-identical polynucleotide sequences; in some  
embodiments, DNA shuffling may involve crossover via nonhomologous  
10                   5                   recombination, such as via cre/lox and/or flp/frt systems, or by oligonucleotide or in  
silico shuffling, or the like, such that recombination need not require substantially  
homologous polynucleotide sequences. Homologous and non-homologous  
15                   recombination formats can be used, and, in some embodiments, can generate  
molecular chimeras and/or molecular hybrids of substantially dissimilar sequences.  
Viral recombination systems, such as template-switching and the like can also be used  
20                   10                   to generate molecular chimeras and recombined genes, or portions thereof. A general  
description of shuffling is provided in commonly-assigned WO98/13487 and  
WO98/13485, both of which are incorporated herein in their entirety by reference; in  
25                   15                   case of any conflicting description of definition between any of the incorporated  
documents and the text of this specification, the present specification provides the  
principal basis for guidance and disclosure of the present invention.

25                   The term "related polynucleotides" means that regions or areas of the  
polynucleotides are identical and regions or areas of the polynucleotides are  
heterologous.

30                   The term "chimeric polynucleotide" means that the polynucleotide  
20                   comprises regions which are wild-type and regions which are mutated. It may also  
mean that the polynucleotide comprises wild-type regions from one polynucleotide  
and wild-type regions from another related polynucleotide.

35                   The term "cleaving" means digesting the polynucleotide with enzymes  
or breaking the polynucleotide (c.g., by chemical or physical means), or generating  
25                   partial length copies of a parent sequence(s) via partial PCR extension, PCR  
stuttering, differential fragment amplification, or other means of producing partial  
40                   length copies of one or more parental sequences.

45                   The term "population" as used herein means a collection of  
30                   components such as polynucleotides, nucleic acid fragments or proteins. A "mixed  
population" means a collection of components which belong to the same family of  
nucleic acids or proteins (i.e. are related) but which differ in their sequence (i.e. are  
not identical) and hence in their biological activity.

5 The term "mutations" means changes in the sequence of a parent  
nucleic acid sequence (e.g., a gene or a microbial genome, transferable element, or  
episome) or changes in the sequence of a parent polypeptide. Such mutations may be  
10 point mutations such as transitions or transversions. The mutations may be deletions,  
5 insertions or duplications.

15 The term "recursive sequence recombination" as used herein refers to a  
method whereby a population of polynucleotide sequences are recombined with each  
other by any suitable recombination means (e.g., sexual PCR, homologous  
10 recombination, site-specific recombination, etc.) to generate a library of sequence-  
recombined species which is then screened or subjected to selection to obtain those  
sequence-recombined species having a desired property; the selected species are then  
20 subjected to at least one additional cycle of recombination with themselves and/or  
with other polynucleotide species and at subsequent selection or screening for the  
desired property.

25 The term "amplification" means that the number of copies of a nucleic  
acid fragment is increased.

30 The term "naturally-occurring" as used herein as applied to an object  
refers to the fact that an object can be found in nature. For example, a polypeptide or  
polynucleotide sequence that is present in an organism that can be isolated from a  
20 source in nature and which has not been intentionally modified by man in the  
laboratory is naturally-occurring. As used herein, laboratory strains and established  
cultivars of plants which may have been selectively bred according to classical  
35 genetics are considered naturally-occurring. As used herein, naturally-occurring  
polynucleotide and polypeptide sequences are those sequences, including natural  
25 variants thereof, which can be found in a source in nature, or which are sufficiently  
similar to known natural sequences that a skilled artisan would recognize that the  
40 sequence could have arisen by natural mutation and recombination processes.

45 As used herein "predetermined" means that the cell type, non-human  
animal, or virus may be selected at the discretion of the practitioner on the basis of a  
30 known phenotype.

50 As used herein, "linked" means in polynucleotide linkage (i.e.,  
phosphodiester linkage). "Unlinked" means not linked to another polynucleotide

sequence; hence, two sequences are unlinked if each sequence has a free 5' terminus and a free 3' terminus.

As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the coding sequence. Operably linked means that the DNA sequences being linked are typically contiguous and, where necessary to join two protein coding regions, contiguous and in reading frame. However, since enhancers generally function when separated from the promoter by several kilobases and intronic sequences may be of variable lengths, some polynucleotide elements may be operably linked but not contiguous. A structural gene which is operably linked to a polynucleotide sequence corresponding to a transcriptional regulatory sequence of an endogenous gene is generally expressed in substantially the same temporal and cell type-specific pattern as is the naturally-occurring gene.

As used herein, the terms "expression cassette" refers to a polynucleotide comprising a promoter sequence and, optionally, an enhancer and/or silencer element(s), operably linked to a structural sequence, such as a cDNA sequence or genomic DNA sequence. In some embodiments, an expression cassette may also include polyadenylation site sequences to ensure polyadenylation of transcripts. When an expression cassette is transferred into a suitable host cell, the structural sequence is transcribed from the expression cassette promoter, and a translatable message is generated, either directly or following appropriate RNA splicing. Typically, an expression cassette comprises: (1) a promoter, such as a CaMV 35S promoter, a NOS promoter or a *rbcS* promoter, or other suitable promoter known in the art, (2) a cloned polynucleotide sequence, such as a cDNA or genomic fragment ligated to the promoter in sense orientation so that transcription from the promoter will produce a RNA that encodes a functional protein, and (3) a polyadenylation sequence. For example and not limitation, an expression cassette of the invention may comprise the cDNA expression cloning vectors, pCD and  $\lambda$ NMT (Okayama H and Berg P (1983) Mol. Cell. Biol. 3: 280; Okayama H and Berg P (1985) Mol. Cell. Biol. 5: 1136, incorporated herein by reference). With reference to

5 expression cassettes which are designed to function in chloroplasts, such as an  
expression cassette encoding a large or small subunit of ADPGPP in a higher plant,  
the expression cassette comprises the sequences necessary to ensure expression in  
chloroplasts or translocation of a nuclear-encoded form translated in the cytoplasm  
10 into the chloroplast. For embodiments wherein the ADPGPP subunits(s) are  
expressed in chloroplasts, typically the subunit encoding sequence is flanked by two  
regions of homology to the plastid genome so as to effect a homologous  
recombination with the chloroplastid genome; often a selectable marker gene is also  
15 present within the flanking plastid DNA sequences to facilitate selection of  
genetically stable transformed chloroplasts in the resultant transplastonic plant cells  
(see Maliga P (1993) TIBTECH 11: 101; Daniell et al. (1998) Nature Biotechnology  
20 16: 346, and references cited therein).

As used herein, the term "transcriptional unit" or "transcriptional  
complex" refers to a polynucleotide sequence that comprises a structural gene  
15 (exons), a cis-acting linked promoter and other cis-acting sequences necessary for  
efficient transcription of the structural sequences, distal regulatory elements necessary  
for appropriate tissue-specific and developmental transcription of the structural  
sequences, and additional cis sequences important for efficient transcription and  
translation (e.g., polyadenylation site, mRNA stability controlling sequences).  
30

20 As used herein, the term "transcription regulatory region" refers to a  
DNA sequence comprising a functional promoter and any associated transcription  
elements (e.g., enhancer, CCAAT box, TATA box, LRE, ethanol-inducible element,  
etc.) that are essential for transcription of a polynucleotide sequence that is operably  
35 linked to the transcription regulatory region.

25 As used herein, the term "xenogeneic" is defined in relation to a  
recipient genome, host cell, or organism and means that an amino acid sequence or  
polynucleotide sequence is not encoded by or present in, respectively, the naturally-  
occurring genome of the recipient genome, host cell, or organism. Xenogenic DNA  
sequences are foreign DNA sequences. Further, a nucleic acid sequence that has been  
45 substantially mutated (e.g., by site directed mutagenesis) is xenogeneic with respect  
30 to the genome from which the sequence was originally derived, if the mutated  
sequence does not naturally occur in the genome.

5                   The term "corresponds to" is used herein to mean that a polynucleotide  
sequence is homologous (i.e., identical) to all or a portion of a reference  
polynucleotide sequence, or that a polypeptide sequence is identical to a reference  
polypeptide sequence. In contradistinction, the term "complementary to" is used  
10                   5                   herein to mean that the complementary sequence is homologous to all or a portion of  
a reference polynucleotide sequence. For illustration, the nucleotide sequence "5'-  
TATAC" corresponds to a reference sequence "5'-TATAC" and is complementary to  
15                   a reference sequence "5'-GTATA".

                  The following terms are used to describe the sequence relationships  
10                   between two or more polynucleotides: "reference sequence", "comparison window",  
"sequence identity", "percentage of sequence identity", and "substantial identity". A  
20                   "reference sequence" is a defined sequence used as a basis for a sequence  
comparison; a reference sequence may be a subset of a larger sequence, for example,  
as a segment of a full-length viral gene or virus genome. Generally, a reference  
25                   15                   sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in  
length, and often at least 50 nucleotides in length. Since two polynucleotides may  
each comprise (1) a sequence (i.e., a portion of the complete polynucleotide  
sequence) that is similar between the two polynucleotides, and (2) a sequence that is  
30                   divergent between the two polynucleotides, sequence comparisons between two (or  
20                   more) polynucleotides are typically performed by comparing sequences of the two  
polynucleotides over a "comparison window" to identify and compare local regions of  
sequence similarity.

35                   A "comparison window", as used herein, refers to a conceptual  
segment of at least 25 contiguous nucleotide positions wherein a polynucleotide  
25                   sequence may be compared to a reference sequence of at least 25 contiguous  
nucleotides and wherein the portion of the polynucleotide sequence in the comparison  
40                   window may comprise additions or deletions (i.e., gaps) of 20 percent or less as  
compared to the reference sequence (which for comparative purposes in this manner  
does not comprise additions or deletions) for optimal alignment of the two sequences.  
45                   30                   Optimal alignment of sequences for aligning a comparison window may be conducted  
by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2:  
482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol.

5            Biol. 48: 443, by the search for similarity method of Pearson and Lipman (1988)  
10            Proc. Natl. Acad. Sci. (U.S.A.) 85: 2444, by computerized implementations of these  
             algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics  
             Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison,  
             WI), or by inspection, and the best alignment (i.e., resulting in the highest percentage  
             of homology over the comparison window) generated by the various methods is  
             selected.

15            The term "sequence identity" means that two polynucleotide sequences  
             are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of  
             comparison. The term "percentage of sequence identity" is calculated by comparing  
             two optimally aligned sequences over the window of comparison, determining the  
20            number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I)  
             occurs in both sequences to yield the number of matched positions, dividing the  
             number of matched positions by the total number of positions in the window of  
             comparison (i.e., the window size), and multiplying the result by 100 to yield the  
25            percentage of sequence identity. The term "substantial identity" as used herein  
             denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide  
             comprises a sequence that has at least 80 percent sequence identity, preferably at least  
30            85 percent identity and often 89 to 95 percent sequence identity, more usually at least  
             99 percent sequence identity as compared to a reference sequence (e.g., a sequence  
             which is a target for recombination) over a comparison window of at least 20  
             nucleotide positions, optionally over a window of at least 30-50 nucleotides, wherein  
35            the percentage of sequence identity is calculated by comparing the reference sequence  
             to the polynucleotide sequence that may include deletions or additions which total 20  
             percent or less of the reference sequence over the window of comparison. The  
40            reference sequence may be a subset of a larger sequence.

             Specific hybridization is defined herein as the formation, by hydrogen  
             bonding or nucleotide (or nucleobase) bases, of hybrids between a probe  
             polynucleotide (e.g., a polynucleotide of the invention and a specific target  
45            polynucleotide, wherein the probe preferentially hybridizes to the specific target such  
             that, for example, a single band corresponding to, e.g., one or more of the RNA  
             species of the gene (or specifically cleaved or processed RNA species) can be



identified on a Northern blot of RNA prepared from a suitable source. Such hybrids may be completely or only partially base-paired. Polynucleotides of the invention which specifically hybridize to viral genome sequences may be prepared on the basis of the sequence data provided herein and available in the patent applications incorporated herein and scientific and patent publications noted above, and according to methods and thermodynamic principles known in the art and described in Sambrooke et al. et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., (1989), Cold Spring Harbor, N.Y.; Berger and Kimmel, Methods in Enzymology, Volume 152, Guide to Molecular Cloning Techniques (1987), Academic Press, Inc., San Diego, CA; Goodspeed et al. (1989) Gene 76: 1; Dunn et al. (1989) J. Biol. Chem. 264: 13057, and Dunn et al. (1988) J. Biol. Chem. 263: 10878, which are each incorporated herein by reference.

"Physiological conditions" as used herein refers to temperature, pH, ionic strength, viscosity, and like biochemical parameters that are compatible with a viable plant organism or agricultural microorganism (e.g., Rhizobium, Agrobacterium, etc.), and/or that typically exist intracellularly in a viable cultured plant cell, particularly conditions existing in the nucleus of said cell. In general, in vitro physiological conditions can comprise 50-200 mM NaCl or KCl, pH 6.5-8.5, 20-45°C and 0.001-10 mM divalent cation (e.g., Mg<sup>++</sup>, Ca<sup>++</sup>); preferably about 150 mM NaCl or KCl, pH 7.2-7.6, 5 mM divalent cation, and often include 0.01-1.0 percent nonspecific protein (e.g., BSA). A non-ionic detergent (Tween, NP-40, Triton X-100) can often be present, usually at about 0.001 to 2%, typically 0.05-0.2% (v/v). Particular aqueous conditions may be selected by the practitioner according to conventional methods. For general guidance, the following buffered aqueous conditions may be applicable: 10-250 mM NaCl, 5-50 mM Tris HCl, pH 5-8, with optional addition of divalent cation(s), metal chelators, nonionic detergents, membrane fractions, antifoam agents, and/or scintillants.

As used herein, the terms "label" or "labeled" refer to incorporation of a detectable marker, e.g., a radiolabeled amino acid or a recoverable label (e.g. biotinyl moieties that can be recovered by avidin or streptavidin). Recoverable labels can include covalently linked polynucleobase sequences that can be recovered by hybridization to a complementary sequence polynucleotide. Various methods of

5 labeling polypeptides, PNAs, and polynucleotides are known in the art and may be  
used. Examples of labels include, but are not limited to, the following: radioisotopes  
(e.g.,  $^3\text{H}$ ,  $^{14}\text{C}$ ,  $^{35}\text{S}$ ,  $^{125}\text{I}$ ,  $^{131}\text{I}$ ), fluorescent or phosphorescent labels (e.g., FITC,  
10 rhodamine, lanthanide phosphors), enzymatic labels (e.g., horseradish peroxidase,  $\beta$ -  
5 galactosidase, luciferase, alkaline phosphatase), biotinyl groups, predetermined  
polypeptide epitopes recognized by a secondary reporter (e.g., leucine zipper pair  
sequences, binding sites for antibodies, transcriptional activator polypeptide, metal  
15 binding domains, epitope tags). In some embodiments, labels are attached by spacer  
arms of various lengths, e.g., to reduce potential steric hindrance.

10 As used herein, the term "statistically significant" means a result (i.e.,  
an assay readout) that generally is at least two standard deviations above or below the  
20 mean of at least three separate determinations of a control assay readout and/or that is  
statistically significant as determined by Student's t-test or other art-accepted measure  
of statistical significance.

15 The term "transcriptional modulation" is used herein to refer to the  
25 capacity to either enhance transcription or inhibit transcription of a structural  
sequence linked in cis; such enhancement or inhibition may be contingent on the  
occurrence of a specific event, such as stimulation with an inducer and/or may only  
30 be manifest in certain cell types.

20 The term "agent" is used herein to denote a chemical compound, a  
mixture of chemical compounds, a biological macromolecule, or an extract made  
35 from biological materials such as bacteria, plants, fungi, or animal cells or tissues.  
Agents are evaluated for potential activity as ADPGPP inhibitors or allosteric  
effectors by inclusion in screening assays described hereinbelow.

25 As used herein, "substantially pure" means an object species is the  
40 predominant species present (i.e., on a molar basis it is more abundant than any other  
individual macromolecular species in the composition), and preferably a substantially  
purified fraction is a composition wherein the object species comprises at least about  
50 percent (on a molar basis) of all macromolecular species present. Generally, a  
45 substantially pure composition will comprise more than about 80 to 90 percent of all  
30 macromolecular species present in the composition. Most preferably, the object  
species is purified to essential homogeneity (contaminant species cannot be detected  
50

in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species.

As used herein, the term "optimized" is used to mean substantially improved in a desired structure or function relative to an initial starting condition, not necessarily the optimal structure or function which could be obtained if all possible combinatorial variants could be made and evaluated, a condition which is typically impractical due to the number of possible combinations and permutations in polynucleotide sequences of significant length (e.g., a complete plant gene or genome).

As used herein, "ADPGPP enzymatic phenotype" means an observable or otherwise detectable phenotype that can be discriminative based on ADPGPP function. For example and not limitation, an ADPGPP enzymatic phenotype can comprise an enzyme  $K_m$  for a substrate,  $K_m$  for an inhibitor ( $K_i$ ),  $K_m$  for an activator ( $K_a$ ),  $V_{max}$ , a turnover rate, an inhibition coefficient ( $K_i$ ), or an observable or otherwise detectable trait that reports ADPGPP function in a cell or clonal progeny thereof, including an adult plant or starch-storing organ thereof, which otherwise lack said trait in the absence of significant ADPGPP function.

As used herein, "complementing subunit" is used principally with reference to ADPGPP enzymes composed of S and L subunits and means an ADPGPP subunit of the opposite type (e.g., an S subunit can be a complementing subunit to an L subunit, and vice versa), wherein when the L and S subunits are present in a cell or *in vitro* reaction vessel under appropriate assay conditions they form a multimer having detectable ADPGPP activity. A complementing subunit can be obtained from the same taxonomic species of organism, or from a xenogenic species. Calibration assays are performed to determine whether a selected first subunit is a complementing subunit with respect to a second subunit; if the first subunit produces a detectable allosteric effect upon the activity, it is deemed for purposes of this disclosure to constitute a complementing subunit.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides methods, reagents, genetically modified plants, plant cells and protoplasts thereof, microbes, and polynucleotides, and compositions relating to the forced evolution of ADPGPP subunit sequences to improve an enzymatic property of a ADPGPP protein. In an aspect, the invention provides a shuffled ADPGPP L subunit which is catalytically active in the presence of a complementing S subunit, which may itself be shuffled, and which exhibits an improved enzymatic profile, such as an increased  $K_m$  for inhibitor, decreased  $K_m$  for activator, and or a decreased  $K_m$  for substrate, increased  $V_{max}$ , or the like.

In a broad aspect, the invention is based, in part, on a method for shuffling polynucleotide sequences that encode a ADPGPP subunit, such as an S subunit gene, L subunit gene, or combinations thereof. The method comprises the step of selecting at least one polynucleotide sequence that encodes an ADPGPP subunit having an enhanced enzymatic phenotype and subjecting said selected polynucleotide sequence to at least one subsequent round of mutagenesis and/or sequence shuffling, and selection for the enhanced phenotype. Preferably, the method is performed recursively on a collection of selected polynucleotide sequences encoding the ADPGPP subunit to iteratively provide polynucleotide sequences encoding ADPGPP subunit species having the desired enhanced enzymatic phenotype.

The invention provides shuffled ADPGPP encoding sequences, wherein said shuffled encoding sequences comprise at least 21 contiguous nucleotides, preferably at least 30 contiguous nucleotides, or more, of a first naturally occurring ADPGPP L gene sequence and at least 21 contiguous nucleotides, preferably at least 30 contiguous nucleotides, or more, of a second naturally occurring ADPGPP L gene sequence, operably linked in reading frame to encode an ADPGPP L subunit which has ADPGPP activity in the presence of a complementing S subunit and/or in the absence of said S subunit, and which has an enhanced enzymatic phenotype. In some variations, it will be possible to use shuffled encoding sequences which have less than 21 contiguous nucleotides identical to a naturally-occurring ADPGPP L gene sequence.

5

10

15

20

25

30

35

40

45

50

55

5

10

15

20

25

30

The invention also provides shuffled ADPGPP encoding sequences, wherein said shuffled encoding sequences comprise at least 21 contiguous nucleotides, preferably at least 30 contiguous nucleotides, or more, of a first naturally occurring ADPGPP S gene sequence and at least 21 contiguous nucleotides, preferably at least 30 contiguous nucleotides, or more, of a second naturally occurring ADPGPP S gene sequence, operably linked in reading frame to encode an ADPGPP S subunit which has a regulatory effect upon a complementing ADPGPP L subunit such that the multimer composed of the shuffled S subunit(s) and the L subunit(s) exhibit ADPGPP activity and wherein the multimer has an enhanced enzymatic phenotype. In some variations, it will be possible to use shuffled encoding sequences which have less than 21 contiguous nucleotides identical to a naturally-occurring ADPGPP gene sequence(s).

The invention provides shuffled ADPGPP S subunit encoding sequences, wherein the shuffled sequences comprise portions of a first parental ADPGPP encoding sequence which comprises at least one mutation in the encoding sequence as compared to the collection of predetermined naturally occurring ADPGPP S subunit sequences.

The invention provides shuffled ADPGPP L subunit encoding sequences, wherein the shuffled sequences comprise portions of a first parental ADPGPP S encoding sequence which comprises at least one mutation in the encoding sequence as compared to the collection of predetermined naturally occurring ADPGPP L subunit sequences.

Generally, the nomenclature used hereafter and the laboratory procedures in cell culture, molecular genetics, virology, and nucleic acid chemistry and hybridization described below are those well known and commonly employed in the art. Standard techniques are used for recombinant nucleic acid methods, polynucleotide synthesis, and microbial culture and transformation (e.g., biolistics, Agrobacterium (Ti plasmid), electroporation, lipofection). Generally enzymatic reactions and purification steps are performed according to the manufacturer's specifications. The techniques and procedures are generally performed according to conventional methods in the art and various general references (see, generally, Sambrook et al. Molecular Cloning: A Laboratory Manual, 2d ed. (1989) Cold

5 Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., which is incorporated  
herein by reference) which are provided throughout this document. The procedures  
therein are believed to be well known in the art and are provided for the convenience  
of the reader. All the information contained therein is incorporated herein by  
10 5 reference.

Oligonucleotides can be synthesized on an Applied Bio Systems  
oligonucleotide synthesizer according to specifications provided by the manufacturer.

15 Methods for PCR amplification are described in the art (PCR  
Technology: Principles and Applications for DNA Amplification ed. HA Erlich,  
10 Freeman Press, New York, NY (1992); PCR Protocols: A Guide to Methods and  
Applications, eds. Innis, Gelfand, Snisky, and White, Academic Press, San Diego,  
20 CA (1990); Mattila et al. (1991) Nucleic Acids Res. 19: 4967; Eckert, K.A. and  
Kunkel, T.A. (1991) PCR Methods and Applications 1: 17; PCR, eds. McPherson,  
Quirk, and Taylor, IRL Press, Oxford; and U.S. Patent 4,683,202, which are  
15 incorporated herein by reference). Leaf PCR is suitable for genotype analysis of  
25 transgene plants.

All sequences referred to herein or equivalents which function in the  
disclosed methods can be retrieved by GenBank database file designation or a  
30 commonly used reference name which is indexed in GenBank or otherwise published  
20 are incorporated herein by reference and are publicly available.

#### Incorporation by Reference of Related Applications

35 The following co-pending patent applications and publications of the present  
inventors and co-workers are incorporated herein by reference for all purposes:  
U.S.S.N. 08/198,431, filed 17 February 1994, PCT/US95/02126 filed 17 February  
25 1995, WO97/20078, U.S. Patent 5,605,793, U.S. Patent 5,358,665, U.S. Patent  
5,270,170, U.S.S.N. 08/425,684 filed 18 April 1995, U.S.S.N. 08/537,874 filed 30  
40 October 1995, U.S.S.N. 08/564,955 filed 30 November 1995, U.S.S.N. 08/621,859  
filed 25 March 1996, PCT/US96/05480 filed 18 April 1996, U.S.S.N. 08/650,400  
filed 20 May 1996, U.S.S.N. 08/675,502 filed 3 July 1996, U.S.S.N. 08/721,824 filed  
45 27 September 1996, U.S.S.N. 08/722,660 filed 27 September 1996, and U.S.S.N.  
30 08/769,062 filed 18 December 1996; WO98/13485 and WO98/13487; and Stemmer  
(1995) Science 270: 1510; Stemmer et al. (1995) Gene 164: 49-53; Stemmer (1995)

5            Bio/Technology 13: 549-553; Stemmer (1994) PNAS 91: 10747-10751; Stemmer  
10            (1994) Nature 370: 389-391; Crameri et al. (1996) Nature Medicine 2: 1-3; Crameri  
15            et al. (1996) Nature Biotechnology 14: 315-319 and commonly assigned U.S. Patent  
20            Application U.S.S.N. 60/107,757 entitled "MODIFIED  
25            PHOSPHOENOLPYRUVATE CARBOXYLASE FOR IMPROVEMENT AND  
30            OPTIMIZATION OF PLANT PHENOTYPES," filed on 10 November 1998  
35            (Attorney Docket Number 018097-029200PC); commonly assigned U.S. Patent  
40            Application U.S.S.N. 60/107,756 and 60/153,093 entitled "MODIFIED RIBULOSE  
45            BISPHOSPHATE CARBOXYLASE/OXYGENASE FOR IMPROVEMENT AND  
50            OPTIMIZATION OF PLANT PHENOTYPES," filed on 10 November 1998 and  
55            September 9, 1999, respectively; and "TRANSFORMATION, SELECTION, AND  
60            SCREENING OF SEQUENCE SHUFFLED POLYNUCLEOTIDES FOR  
65            DEVELOPMENT AND OPTIMIZATION OF PLANT PHENOTYPES" USSN  
70            60/098,528, PCT/US99/19732 and USSN 09/385,833 filed August 31, 1998, August  
75            30, 1999 and August 30, 1999, respectively.

#### Overview

          The invention relates in part to a method for generating novel or  
improved ADPGPP genetic sequences and improved starch production phenotypes  
which do not naturally occur or would be anticipated to occur at a substantial  
frequency in nature. A broad aspect of the method employs recursive nucleotide  
sequence recombination, termed "sequence shuffling", which enables the rapid  
generation of a collection of broadly diverse phenotypes that can be selectively bred  
for a broader range of novel phenotypes or more extreme phenotypes than would  
otherwise occur by natural evolution in the same time period. A basic variation of the  
method is a recursive process comprising: (1) sequence shuffling of a plurality of  
species of a genetic sequence, which species may differ by as little as a single  
nucleotide difference or may be substantially different yet retain sufficient regions of  
sequence similarity or site-specific recombination junction sites to support shuffling  
recombination, (2) selection of the resultant shuffled genetic sequence to isolate or  
enrich a plurality of shuffled genetic sequences having a desired phenotype(s), and (3)  
repeating steps (1) and (2) on the plurality of shuffled genetic sequences having the  
desired phenotype(s) until one or more variant genetic sequences encoding a

5 sufficiently optimized desired phenotype is obtained. In this general manner, the method facilitates the "forced evolution" of a novel or improved genetic sequence to encode a desired ADPGPP enzymatic phenotype which natural selection and evolution has heretofore not generated in the reference agricultural organism.

10 5 Typically, a plurality of ADPGPP genetic sequences are shuffled and selected by the present method. The method can be used with a plurality of alleles, homologs, or cognate genes of a genetic locus, or even with a plurality of genetic sequences from related organisms, and in some instances with unrelated genetic sequences or portions thereof which have recombinogenic portions (either naturally or generated via genetic engineering). Furthermore, the method can be used to evolve a heterologous ADPGPP sequence (e.g., a non-naturally occurring mutant gene, or a subunit from another species) to optimize its function in concert with a complementing subunit, and/or in a particular host cell.

ADPGPP Embodiment - Lowered Km for substrate; Other features

15 25 The invention provides an isolated polynucleotide encoding an enhanced ADPGPP protein having ADPGPP catalytic activity wherein the Km for a substrate (ATP,  $\alpha$ -glucose-1-phosphate (G1P)) is significantly lower than in a protein encoded by a parental polynucleotide encoding a naturally-occurring ADPGPP enzyme. Typically, the Km for substrate will be at least one-half logarithm unit lower than the parental sequence, preferably the Km will be at least one logarithm unit lower, and desirably the Km will be at least two logarithm units lower, or more. The isolated polynucleotide encoding an enhanced ADPGPP protein and in an expressible form can be transferred into a host plant, such as a crop species, wherein suitable expression of the polynucleotide in the host plant will result in improved starch biosynthesis efficiency as compared to the naturally-occurring host plant species, usually under certain conditions. The isolated polynucleotide can encode a single subunit ADPGPP, such as a bacterial form, or may encode a large (L) subunit or small (S) subunit of a multisubunit ADPGPP such as that found in green algae, and higher plants. The isolated polynucleotide can comprise a substantially full-length or full-length coding sequence substantially identical to a naturally occurring S gene and/or an L gene, typically comprising a shuffled L gene or a shuffled S gene, or both.



5 In a variation, the invention provides a polynucleotide comprising: (1)  
a sequence encoding a shuffled ADPGPP L subunit gene operably linked to a  
transcriptional regulatory sequence functional in a host cell, and further linked to (2) a  
10 selectable marker gene which affords a means of selection when expressed in host  
5 cells.

15 In a variation, the invention provides a polynucleotide comprising: (1)  
a sequence encoding a shuffled ADPGPP S subunit gene operably linked to a  
transcriptional regulatory sequence functional in a host cell, and further linked to (2) a  
selectable marker gene which affords a means of selection when expressed in host  
10 cells.

20 In a variation, the invention provides a polynucleotide comprising: (1)  
a sequence encoding a shuffled ADPGPP L subunit gene operably linked to a  
transcriptional regulatory sequence functional in a host cell, (2) a sequence encoding  
a shuffled ADPGPP S subunit gene operably linked to a transcriptional regulatory  
25 sequence functional in the host cell and, optionally, further linked to (3) a selectable  
marker gene which affords a means of selection when expressed in host cells.

30 In a variation, the invention provides an isolated polynucleotide  
encoding an enhanced ADPGPP protein having ADPGPP catalytic activity wherein  
the  $K_m$  for a substrate is significantly lower than a protein encoded by a parental  
20 polynucleotide encoding a naturally-occurring ADPGPP enzyme or subunit. In an  
aspect, the enhanced ADPGPP protein is often an S subunit which is catalytically  
active in the presence of a complementing L subunit. In an aspect, the enhanced  
35 ADPGPP protein is a S subunit which is catalytically active in the absence of a  
complementing L subunit, such as for example, and not limitation, an ADPGPP S  
25 subunit which is at least 90 percent sequence identical to a naturally occurring  
ADPGPP subunit encoded by a genome of a plant or algae.

40 In a variation, the invention provides an isolated polynucleotide  
encoding an enhanced ADPGPP protein having ADPGPP catalytic activity wherein  
the  $K_m$  ( $K_i$ ) for an inhibitor (e.g.,  $P_i$ ) is significantly higher than a protein encoded by  
45 a parental polynucleotide encoding a naturally-occurring ADPGPP enzyme. In such  
embodiments, the concentration of inhibitor required to produce half-maximal  
30

50

55

inhibition of catalysis is typically at least one-half logarithm unit higher than a parental ADPGPP, often at least one log unit or more higher.

In a variation, the invention provides an isolated polynucleotide encoding an enhanced ADPGPP protein having ADPGPP catalytic activity wherein the  $K_m$  for an activator (e.g., PGA, FBP) is significantly lower than in a protein encoded by a parental polynucleotide encoding a naturally-occurring ADPGPP enzyme. In such embodiments, the concentration of activator required to produce half-maximal activation of catalysis is typically at least one-half logarithm unit lower than a parental ADPGPP, often at least one log unit or more lower, in some embodiments at least two log units or more lower. In a variation, the shuffled ADPGPP protein possesses, in the substantial absence of activator, ADPGPP catalytic activity approximately equivalent to or greater than that of a naturally-occurring ADPGPP protein which is maximally stimulated with activator.

The invention provides an enhanced ADPGPP protein having ADPGPP catalytic activity wherein: (1) the  $K_m$  for substrate is significantly lower than in a protein encoded by a parental polynucleotide encoding a naturally-occurring ADPGPP enzyme, and (2) the  $K_m$  for inhibitor is significantly higher than a protein encoded by a parental polynucleotide encoding a naturally-occurring ADPGPP enzyme, and/or (3) the  $K_m$  for activator is significantly lower than in a protein encoded by a parental polynucleotide encoding a naturally-occurring ADPGPP enzyme, and/or (4) the enhanced ADPGPP protein possesses a catalytic activity in the substantial absence of activator and inhibitor which is at least 25 percent or more greater than a naturally-occurring ADPGPP that is maximally stimulated with activator in the substantial absence of inhibitor; often the naturally-occurring ADPGPP used for comparison is an ADPGPP species which has an S subunit polypeptide that has the greatest percentage sequence identity to the shuffled S subunit polypeptide.

In an aspect, the invention provides a polynucleotide sequence encoding an shuffled S subunit of a plant or algal ADPGPP, wherein the shuffled S subunit, either alone and/or when reconstituted with a complementing L subunit, possesses a detectable enzymatic activity wherein: (1) the  $K_m$  for substrate is significantly lower than in an S subunit protein encoded by a parental polynucleotide

5 encoding a naturally-occurring ADPGPP enzyme, (2) the  $K_m$  for an ADPGPP  
inhibitor is significantly higher than an S subunit protein encoded by a parental  
polynucleotide encoding a naturally-occurring ADPGPP enzyme, and/or (3) the  $K_m$   
10 for an ADPGPP activator is significantly lower than a S subunit protein encoded by a  
parental polynucleotide encoding a naturally-occurring ADPGPP enzyme S subunit,  
and/or (4) the  $V_{max}$  for ADPGPP catalytic activity is substantially higher than the  
15  $V_{max}$  for ADPGPP catalytic activity of naturally-occurring ADPGPP under  
equivalent assay conditions (e.g., same concentration(s) of substrates, activators, and  
inhibitors) under at least one assay condition. In a variation, the shuffled S subunit  
10 requires a complementing L subunit for detectable enzymatic activity, or for increased  
enzymatic activity as compared to the activity of the shuffled S subunit in the absence  
of a complementing L subunit. In some embodiments, the shuffled S subunit  
sequences encode proteins that have an altered binding to, or allosteric interaction  
25 with, the complementing L subunit such that the binding constant for an inhibitor or  
activator on the L subunit may be substantially unchanged, however the shuffled S  
subunit, when reconstituted with L subunit, results in formation of an ADPGPP which  
has: (1) reduced sensitivity to inhibitors (e.g.,  $P_i$ ) and/or (2) enhanced sensitivity to  
30 activators (e.g., PGA) or (3) has ADPGPP activity which is insensitive to activator  
and possesses at least one ADPGPP catalytic activity (e.g., substrate  $K_m^{-1}$  or  $V_{max}$ )  
which is at least 25 percent greater than that of a naturally-occurring ADPGPP that is  
maximally stimulated with activator in the substantial absence of inhibitor; often the  
35 naturally-occurring ADPGPP used for comparison is an ADPGPP species which has  
an S subunit polypeptide that has the greatest percentage sequence identity, among  
the collection of then known ADPGPP sequences, to the shuffled S subunit  
25 polypeptide.

40 In an aspect, the invention provides a polynucleotide sequence  
encoding an shuffled L subunit of ADPGPP, wherein the shuffled L subunit possesses  
the property of complexing with an unshuffled, complementing S subunit thereby  
resulting in a multimer (e.g.,  $L_2S_2$ ) having a detectable enzymatic activity wherein: (1)  
45 30 the  $K_m$  for substrate is significantly lower than that of an ADPGPP protein containing  
an L subunit encoded by a parental polynucleotide encoding a naturally-occurring L  
subunit of ADPGPP, (2) the  $V_{max}$  for ADPGPP catalytic activity is significantly

5 higher than that of an ADPGPP protein containing an L subunit encoded by a parental  
polynucleotide encoding a naturally-occurring L subunit of ADPGPP under similar  
assay conditions, and/or (3) the  $K_m$  for activator is significantly lower than that of an  
10 ADPGPP protein containing an L subunit encoded by a parental polynucleotide  
5 encoding a naturally-occurring L subunit of ADPGPP, and/or (4) the  $K_m$  for inhibitor  
is significantly higher than that of an ADPGPP protein containing an L subunit  
encoded by a parental polynucleotide encoding a naturally-occurring L subunit of  
15 ADPGPP, and/or (5) the  $V_{max}$  for ADPGPP catalytic activity is substantially higher  
than the  $V_{max}$  for ADPGPP catalytic activity of naturally-occurring ADPGPP under  
10 equivalent assay conditions (e.g., same concentration(s) of substrates, activators, and  
inhibitors) under at least one assay condition. In some embodiments, the shuffled L  
20 subunit sequences encode proteins that have an altered binding to, or allosteric  
interaction with, the complementing S subunit such that the binding constant for an  
inhibitor or activator on the L subunit may be substantially unchanged, however the  
25 shuffled L subunit, when reconstituted with S subunit, results in formation of an  
ADPGPP which has: (1) reduced sensitivity to inhibitors (e.g.,  $P_i$ ) and/or (2)  
enhanced sensitivity to activators (e.g., PGA) or (3) has ADPGPP activity which is  
insensitive to activator and possesses at least one ADPGPP catalytic activity (e.g.,  
30 substrate  $K_m^{-1}$  or  $V_{max}$ ) which is at least 25 percent greater than that of a naturally-  
occurring ADPGPP that is maximally stimulated with activator in the substantial  
20 absence of inhibitor; often the naturally-occurring ADPGPP used for comparison is  
an ADPGPP species which has an L subunit polypeptide that has the greatest  
35 percentage sequence identity, among the collection of then known ADPGPP  
sequences, to the shuffled L subunit polypeptide. In some embodiments, the binding  
25 constant for an inhibitor, activator, and/or substrate will be at least one-half log unit  
higher or lower than an equivalent naturally occurring ADPGPP of greatest sequence  
40 homology (percent sequence identity) to the shufflant subunit(s).

In an aspect, the invention provides an improved S subunit of an  
ADPGPP, or shufflant thereof, and a polynucleotide encoding same. In some  
45 30 embodiments, the polynucleotide will be operably linked to a transcription regulation  
sequence forming an expression construct, which may be linked to a selectable  
marker gene; for embodiments where it is useful to target a bacterial ADPGPP

5

10

15

20

25

30

35

40

45

50

55

5

10

15

20

25

30

shufflant into plant cell plastids or tuber or other specialized organs where starch synthesis is prominent, a sequence encoding a chloroplast transit peptide (CTP), such as that derived from *Arabidopsis* rbcS gene, is fused in-frame to the shufflant ADPGPP sequence, to ensure delivery of the S subunit to the appropriate compartment/organ. In some embodiments, such a polynucleotide is present as an integrated transgene in a plant chromosome in a format for expression and processing of the S subunit. It can be desirable for such a polynucleotide transgene to be transmissible via germline transmission in a plant; in the case of ADPGPP S gene sequences transferred to a plant or algal cells, it is often accompanied by a selectable marker gene which affords a means to select for progeny which retain the transferred shuffled ADPGPP S gene sequence. In some embodiments, the transferred shuffled ADPGPP S gene sequence is derived by shuffling a pool of parental sequences, at least one of which encodes a bacterial ADPGPP subunit. Often, the transcription control sequences comprise tuber-specific or seed-specific promoters to overcome possible detrimental effects of constitutive expression.

In an aspect, the invention provides an improved S subunit of an ADPGPP, or shufflant thereof, wherein the improved S subunit has at least 80 sequence identity to the polypeptide sequence of a naturally-occurring plant ADPGPP S subunit, and which has an enhanced ADPGPP enzymatic phenotype; and a polynucleotide encoding same. In some embodiments, the polynucleotide will be operably linked to a transcription regulation sequence forming an expression construct, which may be linked to a selectable marker gene. In some embodiments, such a polynucleotide is present as an integrated transgene in a plant chromosome. It can be desirable for such a polynucleotide transgene to be transmissible via germline transmission in a plant. Often, the transcription control sequences comprise tuber-specific or seed-specific promoters to overcome possible detrimental effects of constitutive expression.

In an aspect, the invention provides an improved L subunit of a plant ADPGPP, or shufflant thereof, and a polynucleotide encoding same. In some embodiments, the polynucleotide will be operably linked to a transcription regulation sequence forming an expression construct, which may be linked to a selectable marker gene. In some embodiments, such a polynucleotide is present as an integrated

transgene in a plant chromosome. It can be desirable for such a polynucleotide transgene to be transmissible via germline transmission in a plant.

In an aspect, the invention provides a hybrid S subunit composed of a shufflant comprising a sequence of at least 25 contiguous nucleotides at least 95 percent identical to a plant ADPGPP S gene and a sequence of at least 25 contiguous nucleotides at least 95 percent identical to a bacterial or algal ADPGPP gene, and a polynucleotide encoding same, and typically encoding a substantially full-length ADPGPP S subunit protein, usually comprising at least 90 percent of the coding sequence length, but not necessarily sequence identity, of a naturally occurring ADPGPP S protein. In some embodiments, the polynucleotide will be operably linked to a transcription regulation sequence forming an expression construct, which may be linked to a selectable marker gene. In some embodiments, such a polynucleotide is present as an integrated transgene in a plant chromosome. It can be desirable for such a polynucleotide transgene to be transmissible via germline transmission in a plant.

The invention provides expression constructs, including bacterial plasmids, shuttle vectors, and plant transgenes, wherein the expression construct comprises a transcriptional regulatory sequence functional in plants operably linked to a polynucleotide encoding an enhanced ADPGPP protein subunit. With respect to polynucleotide sequences encoding ADPGPP S subunit proteins, it is generally desirable to express such encoding sequences in plant cells with the expression constructs containing the necessary sequences for appropriate transcription, translation, and processing, which can include translocation to a plastid or other organ compartment. The invention further provides plants and plant germplasm comprising said expression constructs, typically in stably integrated or other replicable form which segregates and can be stably maintained in the host organism, although in some embodiments it is desirable for commercial reasons that the expression sequence not be in the germline of sexually reproducible plants.

The invention provides a method for obtaining an isolated polynucleotide encoding an enhanced ADPGPP protein having ADPGPP catalytic activity wherein the  $K_m$  for substrate is significantly lower than a protein encoded by a parental polynucleotide encoding a naturally-occurring ADPGPP enzyme, the method comprising: (1) recombining sequences of a plurality of parental

5 polynucleotide species encoding at least one ADPGPP sequence under conditions  
suitable for sequence shuffling to form a resultant library of sequence-shuffled  
ADPGPP polynucleotides, (2) transferring said library into a plurality of host cells  
forming a library of transformants wherein sequence-shuffled ADPGPP  
10 5 polynucleotides are expressed, (3) assaying individual or pooled transformants for  
ADPGPP catalytic activity to determine the relative or absolute  $K_m$  for substrate and  
identifying at least one enhanced transformant that expresses a ADPGPP activity  
15 which has a significantly lower  $K_m$  for substrate than the ADPGPP activity encoded  
by the parental sequence(s), (4) recovering the sequence-shuffled ADPGPP  
10 polynucleotide from at least one enhanced transformant. Optionally, the recovered  
sequence-shuffled ADPGPP polynucleotide encoding an enhanced ADPGPP is  
20 recursively shuffled and selected by repeating steps 1 through 4, wherein the  
recovered sequence-shuffled ADPGPP polynucleotide is used as at least one parental  
sequence for subsequent shuffling. If it is desired to obtain a sequence-shuffled  
25 15 ADPGPP encoding a ADPGPP enzyme having an increased  $K_m$  for inhibitor, step 3  
comprises assaying individual or pooled transformants for ADPGPP catalytic activity  
to determine the relative or absolute  $K_m$  for the inhibitor and identifying at least one  
enhanced transformant that expresses a ADPGPP activity which has a significantly  
30 higher  $K_m$  for inhibitor than the ADPGPP activity encoded by the parental  
20 sequence(s). Similarly, if it is desired to obtain a sequence-shuffled ADPGPP  
encoding a ADPGPP enzyme having a decreased  $K_m$  for activator, step 3 comprises  
35 assaying individual or pooled transformants for ADPGPP catalytic activity to  
determine the relative or absolute  $K_m$  for activator, and identifying at least one  
enhanced transformant that expresses an ADPGPP activity which has a significantly  
25 lower  $K_m$  for activator than the ADPGPP activity encoded by the parental  
40 sequence(s).

In an aspect, the method is used to generate sequence-shuffled  
ADPGPP polynucleotides encoding a single subunit ADPGPP which is catalytically  
45 active in the absence of heterologous proteins. For example and not limitation, a  
30 bacterial or algal single subunit ADPGPP gene, such as that from *E. coli* encoded by  
the glgC gene, is shuffled and selected for the desired ADPGPP phenotype (e.g.,  
altered catalytic or regulatory property, or function in a predetermined plant host).

5           The parental single subunit ADPGPP encoding sequence(s) may be shuffled alone or  
in combination with one or more higher plant ADPGPP subunit sequences (L or S),  
preferably those non-bacterial sequences having regions of at least 70 percent  
sequence identity. In an embodiment, a parental ADPGPP encoding sequence  
10           5       employed for generating shufflants comprises an ADPGPP allosteric mutant from *E.*  
*coli* (e.g., SG14, Ala44Thr; CL1136, Arg67Cys; SG5, Pro295Ser; or 618,  
Gly336Asp), *Salmonella typhimurium* (Steiner et al. (1977) *J. Bact.* 129: 246), or the  
15           green algae *Chlamydomonas reinhardtii* (Ball et al. (1991) *Planta* 185: 17).  
Additionally, ADPGPP gene sequences from *Rhodobacter spheroides* or  
10           10       *Rhodospirillum rubrum* can be used. The ADPGPP shufflants are transferred into a  
suitable host cell for expression and selection of the desired ADPGPP phenotype; in  
20           an embodiment, the host cells are *E. coli* strains lacking endogenous ADPGPP  
activity (e.g., LCB618, strains carrying glgC3 mutation or glgC mutation, and the  
like). In an embodiment, the host cells constitutively or inducibly express a  
15           15       complementing ADPGPP subunit (e.g., S or L) to functionally complement the  
shufflant sequences encoding a subunit of a multisubunit form of ADPGPP.

          In an aspect, the ADPGPP gene sequence(s) is/are obtained as an  
isolated polynucleotide and is shuffled by any suitable shuffling method known in the  
30           20       art, such as DNA fragmentation and PCR, error-prone PCR, and the like, preferably  
with one or more additional parental polynucleotides encoding all or a part of another  
ADPGPP species, which may be a single subunit ADPGPP, or one subunit of a  
35           multisubunit ADPGPP, such as a plant L or S subunit. The population of sequence-  
shuffled ADPGPP polynucleotides are each operably linked to an expression  
sequence and transferred into host cells, preferably host cells substantially lacking  
25           25       endogenous ADPGPP activity, such as a deletion strain of *E. coli*, wherein the  
sequence-shuffled ADPGPP polynucleotides are expressed, forming a library of  
40           sequence-shuffled ADPGPP transformants. A sample of individual transformants  
and/or their clonal progeny are isolated into discrete reaction vessels for ADPGPP  
activity assay, or are assayed in situ in certain embodiments. For samples assayed in  
45           30       reaction vessels, aliquots of the samples are separated into a plurality of reaction  
vessels containing an approximately equimolar amount of ADPGPP or total protein,  
and each vessel is assayed for ADPGPP activity in the presence of a predetermined



5 concentration of substrate which ranges from about 0.0001 times the predetermined  
Km for substrate of the ADPGPP encoded by the parental polynucleotide(s) to about  
10,000 times the predetermined Km for substrate of the ADPGPP encoded by the  
parental polynucleotide(s); the plurality of reaction vessels for each shufflant sample  
10 may also contain a fixed or variable concentration of activator and/or inhibitor, or  
neither. From the data generated by assaying the plurality of reaction vessels  
containing aliquots of each transformant, a Km value and/or Vmax is calculated by  
conventional art-known means for the sequence-shuffled ADPGPP of each  
15 transformant; typically the Km and Vmax values for a specific inhibitor or activator  
are determined. Sequence-shuffled polynucleotides encoding ADPGPP proteins that  
have significantly decreased Km and/or Vmax values for substrate, and/or  
20 significantly increased Km values of inhibitor, and/or significantly decreased Km  
values for activator are selected and used as parental sequences for at least one  
additional round of sequence shuffling by any suitable method and selection for  
15 further optimization of the desired ADPGPP phenotype. The shuffling and selection  
process is performed iteratively until sequence shuffled polynucleotides encoding at  
least one ADPGPP enzyme having a desired ADPGPP enzymatic phenotype is  
obtained, or until the optimization to reduce the relevant Km (or increase Vmax) has  
30 plateaued and no further improvement is seen in subsequent rounds of shuffling and  
selection.

In a variation, the sequence-shuffled polynucleotides operably linked  
to an expression sequence is also linked, in polynucleotide linkage, to an expression  
35 cassette encoding a selectable marker gene. Transformants are propagated on a  
selective medium to ensure that transformants which are assayed for ADPGPP  
25 activity contain a sequence-shuffled ADPGPP encoding sequence in expressible form.  
In embodiments wherein a polynucleotide encoding a bacterial ADPGPP are to be  
introduced into host cells which possess plastids, the ADPGPP encoding sequence is  
generally operably linked to a transport sequence to facilitate transport of the  
translated gene product into the plastid. Optionally, a transcriptional regulatory  
45 sequence functional in chloroplasts may be used and the resultant expression cassette  
is transferred into the host cell plastids, such as by biolistics, polyethylene glycol  
(PEG) treatment of protoplasts, or an other suitable method.

5

10

15

20

25

30

35

40

45

50

55

5

10

15

20

25

30

In a variation, the above-described method is modified such that ADPGPP activity is assayed in the presence of varying concentrations of inhibitor and the  $K_m$  for inhibitor is determined. Each vessel containing an aliquot of a transformant is assayed for ADPGPP activity in the presence of a predetermined concentration of inhibitor which ranges from about 0.0001 times the predetermined  $K_m$  for inhibitor of the ADPGPP encoded by the parental polynucleotide(s) to about 10,000 times the predetermined  $K_m$  for inhibitor of the ADPGPP encoded by the parental polynucleotide(s). From the data generated by assaying the plurality of reaction vessels containing aliquots of each transformant, a  $K_m$  value is calculated by conventional art-known means for the sequence-shuffled ADPGPP of each transformant. Sequence-shuffled polynucleotides encoding ADPGPP proteins that have significantly increased  $K_m$  values for inhibitor are selected and used as parental sequences for at least one additional round of sequence shuffling by any suitable method and selection for increased  $K_m$  values for inhibitor. The shuffling and selection process is performed iteratively until sequence shuffled polynucleotides encoding at least one ADPGPP enzyme having a desired  $K_m$  value is obtained, or until the optimization to increase the  $K_m$  has plateaued and no further improvement is seen in subsequent rounds of shuffling and selection.

In a variation, the above-described method is modified such that ADPGPP activity is assayed in the presence of varying concentrations of activator and the  $K_m$  for activator is determined. Each vessel containing an aliquot of a transformant is assayed for ADPGPP activity in the presence of a predetermined concentration of activator which ranges from about 0.0001 times the predetermined  $K_m$  for activator of the ADPGPP encoded by the parental polynucleotide(s) to about 10,000 times the predetermined  $K_m$  for activator of the ADPGPP encoded by the parental polynucleotide(s). From the data generated by assaying the plurality of reaction vessels containing aliquots of each transformant, a  $K_m$  value is calculated by conventional art-known means for the sequence-shuffled ADPGPP of each transformant. Sequence-shuffled polynucleotides encoding ADPGPP proteins that have significantly decreased  $K_m$  values for activator are selected and used as parental sequences for at least one additional round of sequence shuffling by any suitable method and selection for decreased  $K_m$  values for activator. The shuffling and

5 selection process is performed iteratively until sequence shuffled polynucleotides encoding at least one ADPGPP enzyme having a desired  $K_m$  value is obtained, or until the optimization to increase the  $K_m$  has plateaued and no further improvement is seen in subsequent rounds of shuffling and selection.

10 5 In a variation, the method comprises conducting biochemical assays on sample aliquots of transformants to determine ADPGPP enzyme activity so as to establish the ratio of the  $K_m$  for activator to the  $K_m$  for inhibitor for individual transformants. Sequence-shuffled polynucleotides encoding ADPGPP are obtained from transformants exhibiting a decrease in said ratio as compared to the ratio in 15 ADPGPP produced from the parental encoding polynucleotide(s) to provide selected sequence-shuffled ADPGPP polynucleotides which can be used as parental sequences for at least one additional round of sequence shuffling by any suitable method and selection for a decreased ratio of  $K_m(\text{activator})$  to  $K_m(\text{inhibitor})$ . The shuffling and selection process is performed iteratively until sequence shuffled polynucleotides 20 encoding at least one ADPGPP enzyme having a desired  $K_m$  ratio is obtained, or until the optimization to decrease the  $K_m$  ratio has plateaued and no further improvement is seen in subsequent rounds of shuffling and selection.

25 15 In an embodiment of the method, the host cell for transformation with sequence-shuffled polynucleotides encoding ADPGPP is a bacterial mutant which 30 lacks a functional ADPGPP subunit protein, such as *E. coli* glycogen<sup>(-)</sup> mutant or an equivalent. For such mutant host cells, transformants which express ADPGPP activity and permit glycogen synthesis can be readily identified as blue colonies 35 following exposure to iodine vapor, with the degree of blue color serving as a proxy of the degree of ADPGPP activity. In this variation, blue colonies identified after exposure to iodine vapor, or their replicate colonies, are selected and assayed in vitro 40 to determine whether, relative to a parental ADPGPP assayed under equivalent conditions, the  $K_m$  of inhibitor is increased and/or the  $K_m$  for activator is decreased for each shufflant transformant; transformants which exhibit an increased  $K_m(\text{inhibitor})$  and/or a decreased  $K_m(\text{activator})$  are selected and used for at least one 45 subsequent round of sequence shuffling and ADPGPP enzymatic phenotype selection. Often inhibitor-relief shufflants ( $K_m$  for inhibitor is significantly higher than parental) are pooled with each other and reshuffled, as are, separately, activator-relief 50

5 shufflants (Km for activator is significantly decreased compared to parental);  
sometimes inhibitor-relief shufflants and activator-relief shufflants are pooled with  
each other.

10 In an embodiment of the method, the host cell comprises a cell  
5 expressing a complementing subunit of ADPGPP which is capable of interacting with  
an ADPGPP protein encoded by sequence-shuffled polypeptides encoding an  
ADPGPP subunit. For example, if the shuffled polynucleotides encode a large subunit  
15 of ADPGPP, a host cell for the transformation may endogenously encode a small  
subunit of ADPGPP that may interact with a functional large subunit encoded by the  
shuffled polynucleotides. It is often desirable that such host cells lack expression of  
10 the endogenous ADPGPP subunit corresponding to (e.g., cognate to) the type of  
subunit encoded by the shuffled polynucleotides. Mutant cell lines are available in the  
art and novel mutant ADPGPP-deficient cells can be obtained by selecting from a  
pool of mutagenized cells those mutants which have lost detectable ADPGPP activity,  
20 or by homologous gene targeting of ADPGPP L and/or S genes.

25 In an embodiment of the method, polynucleotides encoding naturally-  
occurring ADPGPP protein sequences of a plurality of species of photosynthetic  
prokaryotes and/or algae and/or higher plants are shuffled by a suitable shuffling  
method to generate a shuffled ADPGPP polynucleotide library, wherein each shuffled  
30 ADPGPP encoding sequence is operably linked to an expression sequence, and which  
may optionally comprise a linked selectable marker gene cassette. Said library is  
transformed into a host cell population, such as bacteria which lack endogenous  
ADPGPP activity, to form a transformed host cell library. The transformed host cell  
35 library is propagated on growth medium, which may contain a selection agent to  
ensure retention of a linked selectable marker gene. The transformed host cell library  
25 is subjected to selection by incubating the cells under a graded range of  
concentrations of iodine vapor and selecting blue colonies, preferentially those having  
the deepest coloration of blue. Transformed host cells which are screened for under  
40 the most stringent conditions are isolated individually or in pools, and the sequence-  
shuffled polynucleotide sequences encoding ADPGPP are recovered, and optionally  
45 subjected to at least one subsequent iteration of shuffling and selection on growth  
medium, optionally using lower ranges of iodine vapor pressure (or exposure times)

5 to identify blue colonies. Optionally, or in addition, transformants are assayed for inhibitor-resistant ADPGPP activity and/or high activity ADPGPP in absence of activator. The recovered sequence-shuffled ADPGPP polynucleotide(s) encode(s) an enhanced ADPGPP subunit protein.

10 5 In an embodiment of the method, a host cell comprising a non-photosynthetic bacterium, such as *E. coli*, lacking an endogenous ADPGPP activity, is transformed with an expression cassette encoding the production of a complementing ADPGPP subunit (e.g., S if host cells are to be used with a library of shuffled L genes, and vice-versa), thereby forming a complementing host cell.

15 10 Usually, a linked selectable marker and selection conditions are employed to retain the expression cassette in the complementing host cells and their progeny. ADPGPP encoding sequences are selected by the skilled artisan from publicly available sources. The method further comprises transforming a population of complementing host cells with a library of shuffled ADPGPP-encoding polynucleotides, each

20 15 ADPGPP shufflant polynucleotide encoding a species of a shuffled ADPGPP subunit (S, if the complementing subunit expressed in the host cells is L; L, if the complementing subunit expressed in the host cells is S), then operably linked to a transcriptional control sequence forming a subunit expression cassette, culturing the population of transformed complementing host cells for a suitable incubation period,

25 20 determining the amount of ADPGPP activity in each transformed host cell and its clonal progeny relative to the amount of ADPGPP in untransformed complementing host cells cultured under equivalent conditions, including culture medium, atmosphere, incubation time and temperature, and selecting from said population of transformed complementing host cells and their clonal progeny cells which exhibit

30 25 ADPGPP at statistically significant increased amount relative to said untransformed complementing host cells, and segregating or isolating said selected transformed complementing host cells thereby forming a selected subpopulation of host cells harboring selected shuffled polynucleotides encoding ADPGPP subunit protein species having enhanced catalytic ability; said selected shuffled polynucleotides can

35 30 be recovered and optionally subjected to additional rounds of shuffling and selection for enhanced ADPGPP catalytic or regulatory function to provide one or more optimized shuffled subunit encoding sequences. In a variation, the transformed

5 complementing host cells are segregated in culture vessels, such as a multimicrowell plate, wherein each vessel comprises a subpopulation of species of transformed complementing host cells and their clonal progeny, often consisting of a single species of transformed complementing host cell and its clonal progeny, if any.

10 5 Typically, the expression cassettes encoding the shuffled ADPGPP subunit proteins are linked to a selectable marker gene cassette and selection is applied, typically by selection with an antibiotic in the culture medium, to reduce the prevalence of untransformed cells.

15 The invention provides a plant cell protoplast and clonal progeny thereof containing a sequence-shuffled polynucleotide encoding a ADPGPP subunit which is not encoded by the naturally occurring genome of the plant cell protoplast. 20 The invention also provides a collection of plant cell protoplasts transformed with a library of sequence-shuffled ADPGPP subunit polynucleotides in expressible form. The invention further provides a plant cell protoplast co-transformed with at least two 25 species of library members wherein a first species of library members comprise sequence-shuffled ADPGPP large subunit polynucleotides and a second species of library members comprise sequence-shuffled ADPGPP small subunit polynucleotides. In an embodiment, the subunit polynucleotides are transferred into a plastid 30 compartment for expression and processing, such as by transfer into chloroplasts in a format suitable for expression in the plastid, such as for example and not limitation as a recombinogenic construct for general targeted recombination into a chloroplast chromosome. Alternatively, the subunit proteins encoded by the expression cassettes 35 comprise a chloroplast transit peptide sequence to facilitate transfer of the encoded proteins into the plastid (or other) compartment.

40 25 The invention also provides a regenerated plant containing at least one species of replicable or integrated polynucleotide comprising a sequence-shuffled portion and encoding a ADPGPP subunit polypeptide. The invention provides a method variation wherein at least one round of phenotype selection is performed on regenerated plants derived from protoplasts transformed with sequence-shuffled 45 30 ADPGPP subunit library members. In an embodiment, the phenotype selection comprises a determination, either directly or by proxy, of starch content in a storage

5 tissue (e.g., tuber or seed), or microscopic detection of starch granule size and/or abundance.

10                   5                   The invention provides species-specific ADPGPP shuffling, wherein a transformed plant cell or adult plant or reproductive structure comprises a polynucleotide encoding a shuffled ADPGPP subunit that is at least 95 percent sequence identical to the corresponding ADPGPP subunit encoded by an untransformed naturally-occurring genome of the same taxonomic species of plant cell or adult plant. Typically, the shuffled ADPGPP subunit results from shuffling of one or more alleles encoding the ADPGPP subunit in the taxonomic species genome, optionally including mutagenesis in one or more of the iterative shuffling and selection cycles. The species-specific ADPGPP shuffling may include shuffling a polynucleotide encoding a full-length ADPGPP subunit of a first taxonomic species under conditions whereby ADPGPP subunit sequences of a second taxonomic species (or collection of species) are shuffled in at a low prevalence, such that the resultant population of shufflant polynucleotides contains, on average, shuffled polynucleotides composed of at least about 95 percent sequence encoding the first taxonomic species ADPGPP subunit and less than about 5 percent sequence encoding the second taxonomic species (or collection of species) ADPGPP subunit. The species-specific shufflants are thus highly biased towards identity with the first taxonomic species and shufflants which are selected for the desired ADPGPP phenotype are transferred back into the first taxonomic species for expression and regeneration of adult plants and germplasm. Optionally, selected shufflants are backcrossed against the naturally occurring ADPGPP encoding sequences of the first taxonomic species to remove non-essential sequence alterations and harmonize the final shufflant sequence to the naturally-occurring ADPGPP sequence of the first taxonomic species.

40                   15                   A variation of the method includes adapting a bacterial or algal ADPGPP for optimal function in a plant cell, or adult vegetative plant. This variation comprises recursive shuffling and selection of a library of bacterial or algal ADPGPP encoding sequences in a plant cell of the taxonomic species of plant for which the bacterial or algal ADPGPP is being adapted to function in an adult plant. This variation can include not only selecting for a desired ADPGPP enzymatic phenotype,

5 but also selecting for appropriate function of a operably linked transcriptional control  
sequence, such as a tuber-specific (e.g., patatin promoter) or a seed-specific  
transcriptional control element in conjunction with ADPGPP function. This variation  
can employ host cells which are regenerable post-transformation and selection of  
10 5 adult plants for high starch content storage tissues; recovery of the encoding  
ADPGPP shufflants (and optionally the linked transcriptional control sequences), and  
at least one cycle of recursive shuffling and selection to evolve a bacterial or algal  
ADPGPP, and optionally a transcriptional control sequence, optimized for function in  
15 the desired plant taxonomic species or closely related taxonomic categories.

10 An object of the invention is the production of higher plants which  
express one or more ADPGPP enzyme subunits which confer an enhanced starch  
conversion ratio (or net starch storage amount) to the plants. Although the invention  
is described principally with respect to the use of genetic sequence shuffling to  
generate enhanced ADPGPP coding sequences, the invention also provides for the  
25 15 introduction of ADPGPP coding sequences obtained from organisms having  
ADPGPP with desirable enzymatic phenotypes, such as inhibitor-resistant ADPGPP  
from bacterial mutants, into higher plants. Thus, the invention provides a method  
comprising the step of introducing into a higher plant (e.g., a monocot or dicot) an  
expression cassette encoding an ADPGPP encoded by a genome of a bacterium or  
30 20 algae. Typically, at least a sequence encoding a substantially full-length large  
subunit of the bacterial or algal ADPGPP is transferred. An aspect of the invention  
provides C4 land plants comprising a polynucleotide sequence encoding a bacterial or  
algal ADPGPP composed in an expression cassette suitable for expression in  
chloroplasts of the C4 land plant; optionally an expression cassette encoding an  
25 ADPGPP subunit operably linked to regulatory sequences for expression in the  
nucleus of the C4 plant additionally is transferred into the nucleus of the C4 plant.  
The ADPGPP expression cassette is transferred into the chloroplasts or nucleus of a  
regenerable plant cell (e.g. a protoplast of a C4 plant cell) by art-known  
transformation methods. A C3 plant may be used in place of a C4 plant if desired. A  
45 30 specific embodiment comprises a regenerable protoplast of *Glycine max*, *Nicotiana*  
*tabacum*, or *Zea mays* (or other agricultural crop species amenable to regeneration  
from protoplasts) having a chloroplast or nuclear genome containing an expressible



5 shuffled ADPGPP subunit gene that is obtained from a bacterium or algae, and  
typically is at least 90 percent up to 99 percent sequence identical to an ADPGPP  
subunit gene in the genome of said bacterium or algae, but is mutated in at least one  
10 5 codon as compared to the parental sequence. The invention also provides adult  
plants, cultivars, seeds, vegetative bodies, fruits, germplasm, and reproductive cells  
obtained from regeneration of such transformed protoplasts.

The invention provides a kit for obtaining a polynucleotide encoding a  
15 ADPGPP protein, or subunit thereof, having a predetermined enzymatic phenotype,  
the kit comprising a cell line suitable for forming transformable host cells and a  
10 collection sequence-shuffled polynucleotides formed by *in vitro* sequence shuffling.  
The kit often further comprises a transformation enhancing agent (e.g., lipofection  
20 agent, PEG, etc.) and/or a transformation device (e.g., a biolistics gene gun) and/or a  
plant viral vector which can infect plant cells or protoplasts thereof.

The disclosed method for providing an agricultural organism having an  
25 15 improved ADPGPP enzymatic phenotype by iterative gene shuffling and phenotype  
selection is a pioneering method which enables a broad range of novel and  
advantageous agricultural compositions, methods, kits, uses, plant cultivars, and  
apparatus which will be apparent to those skilled in the art in view of the present  
30 disclosure.

#### 20 ADPGPP

Coding sequences for L and S subunits for various species are  
disclosed in the literature and Genbank, among other public sources, and may be  
35 obtained by cloning, PCR, or from available deposited materials.

ADPGPP subunit shufflants are generated by any suitable shuffling  
25 method from one or more parental sequences, optionally including mutagenesis, and  
the resultant shufflants are introduced into a suitable host cell, typically in the form of  
40 expression cassettes wherein the shuffled polynucleotide sequence encoding the  
ADPGPP subunit is operably linked to a transcriptional regulatory sequence and any  
necessary sequences for ensuring transcription, translation, and processing of the  
45 30 encoded ADPGPP subunit protein. Each such expression cassette or its shuffled  
ADPGPP encoding sequence can be referred to as a "library member" composing a  
library of shuffled ADPGPP subunit sequences. The library is introduced into a  
50

5 population of host cells, such that individual host cells receive substantially one or a  
few species of library member(s), to form a population of shufflant host cells  
expressing a library of shuffled ADPGPP subunit species. The population of  
shufflant host cells is screened so as to isolate or segregate host cells and/or their  
10 5 progeny which express ADPGPP subunit(s) having the desired enhanced phenotype.  
The shuffled ADPGPP subunit encoding sequence(s) is/are recovered from the  
isolated or segregated shufflant host cells, and typically subjected to at least one  
subsequent round of mutagenesis and/or sequence shuffling, introduced into suitable  
15 host cells, and selected for the desired enhanced enzymatic phenotype; this cycle is  
generally performed iteratively until the shufflant host cells express an ADPGPP  
subunit having the desired level or enzymatic phenotype or until the rate of  
20 improvement in the desired enzymatic phenotype produced by shuffling has  
substantially plateaued. The shufflant ADPGPP polynucleotides expressed in the host  
cells following the iterative process of shuffling and selection encode ADPGPP  
subunit specie(s) having the desired enhanced phenotype.  
25

For illustration and not to limit the invention, examples of a desired  
ADPGPP enzymatic phenotype can include increased substrate usage rate at a given  
substrate concentration, decreased inhibition by an ADPGPP inhibitor  
30 (desensitization), increased  $K_m$  for inhibitor (desensitization), increased activation by  
an activator (desensitization), decreased  $K_m$  for activator (desensitization), complete  
20 lack of need for activation (desensitization), decreased ratio of  $K_m$  for activator to  
 $K_m$  for inhibitor, velocity ( $V_{max}$ ) for substrate use, and the like as described herein  
35 and as may be desired by the skilled artisan.

#### Shuffling

25 The following publications describe a variety of recursive  
recombination procedures and/or methods which can be incorporated into such  
procedures, e.g., for shuffling of ADPGPP genes and gene fragments as herein:  
Stemmer, et al., (1999) "Molecular breeding of viruses for targeting and other clinical  
properties. Tumor Targeting" 4:1-4; Nasset al. (1999) "DNA Shuffling of  
45 30 subgenomic sequences of subtilisin" Nature Biotechnology 17:893-896; Chang et al.  
(1999) "Evolution of a cytokine using DNA family shuffling" Nature Biotechnology  
17:793-797; Minshull and Stemmer (1999) "Protein evolution by molecular breeding"

5                    Current Opinion in Chemical Biology 3:284-290; Christians et al. (1999) "Directed  
evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling"  
10                    Nature Biotechnology 17:259-264; Crameriet al. (1998) "DNA shuffling of a family  
of genes from diverse species accelerates directed evolution" Nature 391:288-291;  
15                    Cramer et al. (1997) "Molecular evolution of an arsenate detoxification pathway by  
DNA shuffling," Nature Biotechnology 15:436-438; Zhang et al. (1997) "Directed  
evolution of an effective fucosidase from a galactosidase by DNA shuffling and  
screening" Proceedings of the National Academy of Sciences, U.S.A. 94:4504-4509;  
20                    Patten et al. (1997) "Applications of DNA Shuffling to Pharmaceuticals and  
Vaccines" Current Opinion in Biotechnology 8:724-733; Cramer et al. (1996)  
"Construction and evolution of antibody-phage libraries by DNA shuffling" Nature  
25                    Medicine 2:100-103; Cramer et al. (1996) "Improved green fluorescent protein by  
molecular evolution using DNA shuffling" Nature Biotechnology 14:315-319; Gates  
et al. (1996) "Affinity selective isolation of ligands from peptide libraries through  
display on a lac repressor 'headpiece dimer'" Journal of Molecular Biology 255:373-  
386; Stemmer (1996) "Sexual PCR and Assembly PCR" In: The Encyclopedia of  
30                    Molecular Biology. VCH Publishers, New York. pp.447-457; Cramer and Stemmer  
(1995) "Combinatorial multiple cassette mutagenesis creates all the permutations of  
mutant and wildtype cassettes" BioTechniques 18:194-195; Stemmer et al., (1995)  
35                    "Single-step assembly of a gene and entire plasmid from large numbers of  
oligodeoxyribonucleotides" Gene, 164:49-53; Stemmer (1995) "The Evolution of  
Molecular Computation" Science 270: 1510; Stemmer (1995) "Searching Sequence  
Space" Bio/Technology 13:549-553; Stemmer (1994) "Rapid evolution of a protein in  
vitro by DNA shuffling" Nature 370:389-391; and Stemmer (1994) "DNA shuffling  
by random fragmentation and reassembly: In vitro recombination for molecular  
40                    evolution." Proceedings of the National Academy of Sciences, U.S.A. 91:10747-  
10751.

                    Additional details regarding DNA shuffling methods are found in U.S.  
Patents by the inventors and their co-workers, including: United States Patent  
45                    5,605,793 to Stemmer (February 25, 1997), "METHODS FOR IN VITRO  
RECOMBINATION;" United States Patent 5,811,238 to Stemmer et al. (September  
22, 1998) "METHODS FOR GENERATING POLYNUCLEOTIDES HAVING

DESIRED CHARACTERISTICS BY ITERATIVE SELECTION AND

RECOMBINATION;" United States Patent 5,830,721 to Stemmer et al. (November 3, 1998), "DNA MUTAGENESIS BY RANDOM FRAGMENTATION AND REASSEMBLY;" United States Patent 5,834,252 to Stemmer, et al. (November 10, 1998) "END-COMPLEMENTARY POLYMERASE REACTION," and United States Patent 5,837,458 to Minshull, et al. (November 17, 1998), "METHODS AND COMPOSITIONS FOR CELLULAR AND METABOLIC ENGINEERING."

In addition, details and formats for DNA shuffling are found in a variety of PCT and foreign patent application publications, including: Stemmer and Crameri, "DNA MUTAGENESIS BY RANDOM FRAGMENTATION AND REASSEMBLY" WO 95/22625; Stemmer and Lipschutz "END COMPLEMENTARY POLYMERASE CHAIN REACTION" WO 96/33207; Stemmer and Crameri "METHODS FOR GENERATING POLYNUCLEOTIDES HAVING DESIRED CHARACTERISTICS BY ITERATIVE SELECTION AND RECOMBINATION" WO 97/0078; Minshull and Stemmer, "METHODS AND COMPOSITIONS FOR CELLULAR AND METABOLIC ENGINEERING" WO 97/35966; Punnonen et al. "TARGETING OF GENETIC VACCINE VECTORS" WO 99/41402; Punnonen et al. "ANTIGEN LIBRARY IMMUNIZATION" WO 99/41383; Punnonen et al. "GENETIC VACCINE VECTOR ENGINEERING" WO 99/41369; Punnonen et al. OPTIMIZATION OF IMMUNOMODULATORY PROPERTIES OF GENETIC VACCINES WO 99/41368; Stemmer and Crameri, "DNA MUTAGENESIS BY RANDOM FRAGMENTATION AND REASSEMBLY" EP 0934999; Stemmer "EVOLVING CELLULAR DNA UPTAKE BY RECURSIVE SEQUENCE RECOMBINATION" EP 0932670; Stemmer et al., "MODIFICATION OF VIRUS TROPISM AND HOST RANGE BY VIRAL GENOME SHUFFLING" WO 99/23107; Apt et al., "HUMAN PAPILLOMAVIRUS VECTORS" WO 99/21979; Del Cardayre et al. "EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION" WO 98/1837; Patten and Stemmer, "METHODS AND COMPOSITIONS FOR POLYPEPTIDE ENGINEERING" WO 98/27230; Stemmer et al., and "METHODS FOR OPTIMIZATION OF GENE THERAPY BY RECURSIVE SEQUENCE SHUFFLING AND SELECTION" WO9813487.

5 Certain U.S. Applications provide additional details regarding DNA  
shuffling and related techniques, including "SHUFFLING OF CODON ALTERED  
GENES" by Patten et al. filed September 29, 1998, (USSN 60/102,362), January 29,  
10 1999 (USSN 60/117,729), and September 28, 1999, USSN09/407,800 (Attorney  
5 Docket Number 20-28520US/PCT); "EVOLUTION OF WHOLE CELLS AND  
ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION", by del Cardyre  
et al. filed July 15, 1998 (USSN 09/166,188), and July 15, 1999 (USSN 09/354,922);  
15 "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" by  
Cramer et al., filed February 5, 1999 (USSN 60/118,813) and filed June 24, 1999  
10 (USSN 60/141,049) and filed September 28, 1999 (USSN 09/408,392, Attorney  
Docket Number 02-29620US); and "USE OF CODON-BASED  
20 OLIGONUCLEOTIDE SYNTHESIS FOR SYNTHETIC SHUFFLING" by Welch et  
al., filed September 28, 1999 (USSN 09/408,393, Attorney Docket Number 02-  
010070US); and "METHODS FOR MAKING CHARACTER STRINGS,  
25 POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED  
CHARACTERISTICS" by Selifonov and Stemmer, filed February 5, 1999 (USSN  
60/118854) and "METHODS FOR MAKING CHARACTER STRINGS,  
30 POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED  
CHARACTERISTICS" by Selifonov et al. filed October 12, 1999 (USSN  
20 09/416375).

As review of the foregoing publications, patents, published  
35 applications and U.S. patent applications reveals, recursive recombination and  
selection of nucleic acids to provide new nucleic acids with desired properties can be  
carried out by a number of established methods. Any of these methods can be  
25 adapted to the present invention to evolve ADPGPP coding nucleic acids or  
homologues to produce new enzymes with improved properties. Both the methods of  
40 making such enzymes and the enzymes or enzyme coding libraries produced by these  
methods are a feature of the invention.

In brief, at least 5 different general classes of recombination methods  
45 are applicable to the present invention. First, nucleic acids can be recombined in vitro  
30 by any of a variety of techniques discussed in the references above, including e.g.,  
DNase digestion of nucleic acids to be recombined followed by ligation and/or PCR  
50

5 reassembly of the nucleic acids. Second, nucleic acids can be recursively recombined  
in vivo, e.g., by allowing recombination to occur between nucleic acids in cells.  
Third, whole cell genome recombination methods can be used in which whole  
10 genomes of cells are recombined, optionally including spiking of the genomic or  
5 chloroplast recombination mixtures with desired library components such as  
ADPGPP encoding nucleic acids. Fourth, synthetic recombination methods can be  
used, in which oligonucleotides corresponding to different ADPGPP homologues are  
15 synthesized and reassembled in PCR or ligation reactions which include  
oligonucleotides which correspond to more than one parental nucleic acid, thereby  
10 generating new recombined nucleic acids. Oligonucleotides can be made by standard  
nucleotide addition methods, or can be made, e.g., by tri-nucleotide synthetic  
20 approaches. Fifth, in silico methods of recombination can be effected in which  
genetic algorithms are used in a computer to recombine sequence strings which  
correspond to ADPGPP homologues. The resulting recombined sequence strings are  
25 optionally converted into nucleic acids by synthesis of nucleic acids which  
correspond to the recombined sequences, e.g., in concert with oligonucleotide  
synthesis/ gene reassembly techniques. Any of the preceding general recombination  
formats can be practiced in a reiterative fashion to generate a more diverse set of  
30 recombinant nucleic acids.

20 A basic format of the method, termed sequence shuffling (or simply  
"shuffling"), in broad application, consists of a method for generating a selected  
35 polynucleotide sequence or population of selected polynucleotide sequences, typically  
in the form of amplified and/or cloned polynucleotides, whereby the selected  
polynucleotide sequence(s) possess or encode a desired phenotypic characteristic  
25 (e.g., encode a polypeptide, promote transcription of linked polynucleotides, modify  
transformation efficiency, bind a protein, and the like) which can be selected for. One  
40 method of identifying polypeptides that possess a desired structure or functional  
property, such as encoding a desired enzymatic function(s) (e.g., an enhanced  
ADPGPP, a herbicide catabolizing enzyme, an optimized plant biosynthetic pathway),  
45 involves the screening of a large library of polynucleotides for individual library  
30 members which possess or encode the desired structure or functional property  
conferred by the polynucleotide sequence.

In a general aspect, the invention provides a method, termed "sequence shuffling", for generating libraries of recombinant polynucleotides having a desired ADPGPP enzyme characteristic which can be selected or screened for. Libraries of recombinant polynucleotides are generated from a population of related-sequence polynucleotides which comprise sequence regions which have substantial sequence identity and can be homologous recombined in vitro or in vivo. In the method, at least two species of the related-sequence polynucleotides are combined in a recombination system suitable for generating sequence-recombined polynucleotides, wherein said sequence-recombined polynucleotides comprise a portion of at least one first species of a related-sequence polynucleotide with at least one adjacent portion of at least one second species of a related-sequence polynucleotide. Recombination systems suitable for generating sequence-recombined polynucleotides can be either: (1) in vitro systems for homologous recombination or sequence shuffling via amplification or other formats described herein, or (2) in vivo systems for homologous recombination or site-specific recombination as described herein. The population of sequence-recombined polynucleotides comprises a subpopulation of polynucleotides which possess desired or advantageous characteristics and which can be selected by a suitable selection or screening method. The selected sequence-recombined polynucleotides, which are typically related-sequence polynucleotides, can then be subjected to at least one recursive cycle wherein at least one selected sequence-recombined polynucleotide is combined with at least one distinct species of related-sequence polynucleotide (which may itself be a selected sequence-recombined polynucleotide) in a recombination system suitable for generating sequence-recombined polynucleotides, such that additional generations of sequence-recombined polynucleotide sequences are generated from the selected sequence-recombined polynucleotides obtained by the selection or screening method employed. In this manner, recursive sequence recombination generates library members which are sequence-recombined polynucleotides possessing desired characteristics. Such characteristics can be any property or attribute capable of being selected for or detected in a screening system, and may include properties of: an encoded protein, a transcriptional element, a sequence controlling transcription, RNA processing, RNA stability, chromatin conformation, translation, or other expression property of a gene

5 or transgene, a replicative element, a protein-binding element, or the like, such as any feature which confers a selectable or detectable property.

10 Nucleic acid sequence shuffling is a method for recursive *in vitro* or *in vivo* homologous or nonhomologous recombination of pools of nucleic acid fragments or polynucleotides (e.g., genes from agricultural organisms or portions thereof).  
15 Mixtures of related nucleic acid sequences or polynucleotides are randomly or pseudo randomly fragmented, and reassembled to yield a library or mixed population of recombinant nucleic acid molecules or polynucleotides.

20 The present invention is directed to a method for generating a selected polynucleotide sequence (e.g., a plant ADPGPP gene or microbe ADPGPP gene, or combinations thereof) or population of selected polynucleotide sequences, typically in the form of amplified and/or cloned polynucleotides, whereby the selected polynucleotide sequence(s) possess a desired phenotypic characteristic of ADPGPP enzymes or subunits thereof which can be selected for, and whereby the selected  
25 polynucleotide sequences are genetic sequences having a desired functionality and/or conferring a desired phenotypic property to an agricultural organism in which the polynucleotide has been transferred into.

30 In a general aspect, the invention provides a method, called "sequence shuffling", for generating libraries of recombinant polynucleotides having a subpopulation of library members which encode an enhanced or improved ADPGPP L or S protein. Libraries of recombinant polynucleotides are generated from a population of related-sequence ADPGPP polynucleotides which comprise sequence regions which have substantial sequence identity and can be homologous recombined  
35 *in vitro* or *in vivo*. In the method, at least two species of the related-sequence ADPGPP polynucleotides are combined in a recombination system suitable for generating sequence-recombined polynucleotides, wherein said sequence-recombined polynucleotides comprise a portion of at least one first species of a related-sequence ADPGPP polynucleotide with at least one adjacent portion of at least one second species of a related-sequence ADPGPP polynucleotide. Recombination systems  
40 suitable for generating sequence-recombined polynucleotides can be either: (1) *in vitro* systems for homologous recombination or sequence shuffling via amplification or other formats described herein, or (2) *in vivo* systems for homologous



5 recombination or site-specific recombination as described herein, or template-  
switching of a retroviral genome replication event. The population of sequence-  
recombined polynucleotides comprises a subpopulation of ADPGPP polynucleotides  
which possess desired or advantageous enzymatic characteristics and which can be  
10 5 selected by a suitable selection or screening method. The selected sequence-  
recombined ADPGPP polynucleotides, which are typically related-sequence  
polynucleotides, can then be subjected to at least one recursive cycle wherein at least  
one selected sequence-recombined ADPGPP polynucleotide is combined with at least  
15 one distinct species of related-sequence ADPGPP polynucleotide (which may itself be  
a selected sequence-recombined polynucleotide) in a recombination system suitable  
for generating sequence-recombined ADPGPP polynucleotides, such that additional  
generations of sequence-recombined polynucleotide sequences are generated from the  
selected sequence-recombined polynucleotides obtained by the selection or screening  
method employed. In this manner, recursive sequence recombination generates  
20 15 library members which are sequence-recombined polynucleotides possessing desired  
ADPGPP enzymatic characteristics. Such characteristics can be any property or  
attribute capable of being selected for or detected in a screening system.

Screening/selection produces a subpopulation of genetic sequences (or  
30 cells) expressing recombinant forms of ADPGPP subunit gene(s) that have evolved  
toward acquisition of a desired enzymatic property. These recombinant forms can  
then be subjected to further rounds of recombination and screening/selection in any  
order. For example, a second round of screening/selection can be performed  
35 analogous to the first resulting in greater enrichment for genes having evolved toward  
acquisition of the desired enzymatic property. Optionally, the stringency of selection  
can be increased between rounds (e.g., if selecting for drug resistance, the  
25 concentration of drug in the media can be increased). Further rounds of  
recombination can also be performed by an analogous strategy to the first round  
generating further recombinant forms of the gene(s) or genome(s). Alternatively,  
further rounds of recombination can be performed by any of the other molecular  
40 breeding formats discussed. Eventually, a recombinant form of the ADPGPP subunit  
gene(s) is generated that has fully acquired the desired enzymatic property.

5 In an embodiment, the first plurality of selected library members is  
fragmented and homologous recombined by PCR in vitro. Fragment generation is by  
nuclease digestion, partial extension PCR amplification, PCR stuttering, or other  
10 suitable fragmenting means, such as described herein and in WO95/22625 published  
24 August 1995, and in commonly owned U.S.S.N. U.S.S.N. 08/621,859 filed 25  
March 1996, PCT/US96/05480 filed 18 April 1996, which are incorporated herein by  
reference). Stuttering is fragmentation by incomplete polymerase extension of  
15 templates. A recombination format based on very short PCR extension times can be  
employed to create partial PCR products, which continue to extend off a different  
10 template in the next (and subsequent) cycle(s), and effect de facto fragmentation.  
Template-switching and other formats which accomplish sequence shuffling between  
20 a plurality of sequence-related polynucleotides can be used. Such alternative formats  
will be apparent to those skilled in the art.

In an embodiment, the first plurality of selected  
25 library members is fragmented in vitro, the resultant fragments transferred into a host  
cell or organism and homologous recombined to form shuffled library members in  
vivo.

In an embodiment, the first plurality of selected library members is  
30 cloned or amplified on episomally replicable vectors, a multiplicity of said vectors is  
transferred into a cell and homologous recombined to form shuffled library members  
20 in vivo.

In an embodiment, the first plurality of selected library members is not  
35 fragmented, but is cloned or amplified on an episomally replicable vector as a direct  
repeat or indirect (or inverted) repeat, which each repeat comprising a distinct species  
25 of selected library member sequence, said vector is transferred into a cell and  
40 homologous recombined by intra-vector or inter-vector recombination to form  
shuffled library members in vivo.

In an embodiment, combinations of in vitro and in vivo shuffling are  
45 provided to enhance combinatorial diversity. The recombination cycles (in vitro or in  
30 vivo) can be performed in any order desired by the practitioner.

In one embodiment, the first plurality of selected library members is  
50 fragmented and homologous recombined by PCR in vitro. Fragment generation is by

nuclease digestion, partial extension PCR amplification, PCR stuttering, or other suitable fragmenting means, such as described herein and in the documents incorporated herein by reference. Stuttering is fragmentation by incomplete polymerase extension of templates.

In one embodiment, the first plurality of selected library members is fragmented in vitro, the resultant fragments transferred into a host cell or organism and homologous recombined to form shuffled library members in vivo. In an aspect, the host cell is a plant cell which has been engineered to contain enhanced recombination systems, such as an enhanced system for general homologous recombination (e.g., a plant expressing a recA protein or a plant recombinase from a transgene or plant virus) or a site-specific recombination system (e.g., a cre/LOX or frt/FLP system encoded on a transgene or plant virus).

In one embodiment, the first plurality of selected library members is cloned or amplified on episomally replicable vectors, a multiplicity of said vectors is transferred into a cell and homologous recombined to form shuffled library members in vivo in a plant cell, algae cell, or bacterial cell. Other cell types may be used, if desired.

In one embodiment, the first plurality of selected library members is not fragmented, but is cloned or amplified on an episomally replicable vector as a direct repeat or indirect (or inverted) repeat, which each repeat comprising a distinct species of selected library member sequence, said vector is transferred into a cell and homologous recombined by intra-vector or inter-vector recombination to form shuffled library members in vivo in a plant cell, algae cell, or microorganism.

In an embodiment, combinations of in vitro and in vivo shuffling are provided to enhance combinatorial diversity.

At least two additional related specific formats are useful in the practice of the present invention. The first, referred to as "in silico" shuffling utilizes computer algorithms to perform "virtual" shuffling using genetic operators in a computer. As applied to the present invention, ADPGPP nucleic acid (or protein) sequence strings are recombined in a computer system and desirable products are made, e.g., by reassembly PCR or ligation of synthetic oligonucleotides, or other available techniques. In silico shuffling is described in detail in Selifonov and

5 Stemmer in "METHODS FOR MAKING CHARACTER STRINGS,  
POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED  
CHARACTERISTICS" filed 02/05/1999, USSN 60/118854 and "METHODS FOR  
10 MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES  
5 HAVING DESIRED CHARACTERISTICS" by Selifonov et al. filed October 12,  
1999 (USSN 09/416375). In brief, genetic operators (algorithms which represent  
given genetic events such as point mutations, recombination of two strands of  
15 homologous nucleic acids, etc.) are used to model recombinational or mutational  
events which can occur in one or more nucleic acid, e.g., by aligning nucleic acid  
10 sequence strings (using standard alignment software, or by manual inspection and  
alignment) and predicting recombinational outcomes based upon selected genetic  
20 algorithms (mutation, recombination, etc.). The predicted recombinational outcomes  
are used to produce corresponding molecules, e.g., by oligonucleotide synthesis and  
reassembly PCR. As applied to the present invention, ADPGPP nucleic acids are  
15 aligned and recombined in silico, using any desired genetic operator, to produce  
25 character strings which are then generated synthetically for subsequent screening.

The second useful format is referred to as "oligonucleotide mediated  
shuffling" in which oligonucleotides corresponding to a family of related homologous  
30 nucleic acids (e.g., as applied to the present invention, families of homologous  
20 ADGPP variants of a nucleic acid) which are recombined to produce selectable  
nucleic acids. This format is described in detail in Crameri et al.

35 "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" filed  
February 5, 1999, USSN 60/118,813, Crameri et al. "OLIGONUCLEOTIDE  
MEDIATED NUCLEIC ACID RECOMBINATION" filed June 24, 1999, USSN  
25 60/141,049; Crameri et al. "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID  
40 RECOMBINATION" filed September 28, 1999 (USSN 09/408,392, Attorney Docket  
Number 02-29620US); and "USE OF CODON-BASED OLIGONUCLEOTIDE  
SYNTHESIS FOR SYNTHETIC SHUFFLING" by Welch et al., filed September 28,  
45 1999 (USSN 09/408,393, Attorney Docket Number 02-010070US). In brief, selected  
30 oligonucleotides corresponding to multiple homologous parental nucleic acids are  
synthesized, ligated and elongated (typically in a recursive format), typically either in  
a polymerase or ligase-mediated elongation reaction, to produce full-length ADPGPP

5 nucleic acids. The technique can be used to recombine homologous or even non-homologous ADPGPP nucleic acid sequences.

10 One advantage of oligonucleotide-mediated recombination is the ability to recombine homologous nucleic acids with low sequence similarity, or even non-homologous nucleic acids. In these low-homology oligonucleotide shuffling methods, one or more set of fragmented nucleic acids (e.g., oligonucleotides 5 corresponding to multiple ADPGPP nucleic acids) are recombined, e.g., with a set of crossover family diversity oligonucleotides. Each of these crossover oligonucleotides have a plurality of sequence diversity domains corresponding to a plurality of 10 sequence diversity domains from homologous or non-homologous nucleic acids with low sequence similarity. The fragmented oligonucleotides, which are derived by comparison to one or more homologous or non-homologous nucleic acids, can 20 hybridize to one or more region of the crossover oligos, facilitating recombination.

When recombining homologous nucleic acids, sets of overlapping 15 family gene shuffling oligonucleotides (which are derived by comparison of homologous nucleic acids, by synthesis of corresponding oligonucleotides) are hybridized and elongated (e.g., by reassembly PCR or ligation), providing a population of recombined nucleic acids, which can be selected for a desired trait or 25 property. The set of overlapping family shuffling gene oligonucleotides includes a plurality of oligonucleotide member types which have consensus region subsequences derived from a plurality of homologous target nucleic acids. 30

Typically, as applied to the present invention, family gene shuffling 35 oligonucleotides which include one or more ADPGPP nucleic acid(s) are provided by aligning homologous nucleic acid sequences to select conserved regions of sequence identity and regions of sequence diversity. A plurality of family gene shuffling 25 oligonucleotides are synthesized (serially or in parallel) which correspond to at least one region of sequence diversity. 40

Sets of fragments, or subsets of fragments used in oligonucleotide shuffling approaches can be provided by cleaving one or more homologous nucleic 45 acids (e.g., with a DNase), or, more commonly, by synthesizing a set of oligonucleotides corresponding to a plurality of regions of at least one nucleic acid 30 (typically oligonucleotides corresponding to a full-length nucleic acid are provided as 50

members of a set of nucleic acid fragments). In the shuffling procedures herein, these cleavage fragments can be used in conjunction with family gene shuffling oligonucleotides, e.g., in one or more recombination reaction to produce recombinant ADPGPP nucleic acid(s).

One final synthetic variant worth noting is found in "SHUFFLING OF CODON ALTERED GENES" by Patten et al. filed September 29, 1998, (USSN 60/102,362), January 29, 1999 (USSN 60/117,729), and September 28, 1999, PCT/US99/22588 (Attorney Docket Number 20-28520US/PCT). As noted in detail in this set of related applications, one way of generating diversity in a set of nucleic acids to be shuffled (i.e., as applied to the present invention, ADPGPP nucleic acids), is to provide codon-altered nucleic acids which can be shuffled to provide access to sequence space not present in naturally occurring sequences. In brief, by synthesizing nucleic acids in which the codons which encode polypeptides are altered, it is possible to access a completely different mutational spectrum upon subsequent mutation of the nucleic acid. This increases the sequence diversity of the starting nucleic acids for shuffling protocols, which alters the rate and results of forced evolution procedures. Codon modification procedures can be used to modify any ADGPP nucleic acid or shuffled nucleic acid, e.g., prior to performing DNA shuffling.

In brief, oligonucleotide sets comprising codon variations are synthesized and reassembled into full-length nucleic acids. The full length nucleic acids can themselves be shuffled (e.g., where the oligonucleotides to be reassembled provide sequence diversity at selected sites), and/or the full-length sequences can be shuffled by any available procedure to produce diverse sets of ADGPP nucleic acids.

#### Improved Plants

Without reciting the various generalized formats of polynucleotide sequence shuffling and selection described previously or hereinbelow, which will be referred to herein by the shorthand "shuffling", the present invention provides methods, compositions, and uses related to creating novel or improved plants, plant cells, algal cells, soil microbes, plant pathogens, commensal microbes, or other plant-related organisms having art-recognized importance to the agricultural, horticultural, and argonomic areas (collectively, "agricultural organisms").

For example, agronomically and horticulturally important plant species can be transduced. Such species include, but are not restricted to, members of the families: *Graminae* (including corn, rye, triticale, barley, millet, rice, wheat, oats, etc.); *Leguminosae* (including pea, beans, lentil, peanut, yam bean, cowpeas, velvet beans, soybean, clover, alfalfa, lupine, vetch, lotus, sweet clover, wisteria, and sweetpea); *Compositae* (the largest family of vascular plants, including at least 1,000 genera, including important commercial crops such as sunflower) and *Rosaciae* (including raspberry, apricot, almond, peach, rose, etc.), as well as nut plants (including, walnut, pecan, hazelnut, etc.).

Additionally, preferred targets include plants from the genera: *Agrostis*, *Allium*, *Antirrhinum*, *Apium*, *Arachis*, *Asparagus*, *Atropa*, *Avena* (e.g., oats), *Bambusa*, *Brassica*, *Bromus*, *Browaalia*, *Camellia*, *Cannabis*, *Capsicum*, *Cicer*, *Chenopodium*, *Chichorium*, *Citrus*, *Coffea*, *Coix*, *Cucumis*, *Curcubita*, *Cynodon*, *Dactylis*, *Datura*, *Daucus*, *Digitalis*, *Dioscorea*, *Elaeis*, *Eleusine*, *Festuca*, *Fragaria*, *Geranium*, *Glycine*, *Helianthus*, *Heterocallis*, *Hevea*, *Hordeum* (e.g., barley), *Hyoscyamus*, *Ipomoea*, *Lactuca*, *Lens*, *Lilium*, *Linum*, *Lolium*, *Lotus*, *Lycopersicon*, *Majorana*, *Malus*, *Mangifera*, *Manihot*, *Medicago*, *Nemesia*, *Nicotiana*, *Onobrychis*, *Oryza* (e.g., rice), *Panicum*, *Pelargonium*, *Pennisetum* (e.g., millet), *Petunia*, *Pisum*, *Phaseolus*, *Phleum*, *Poa*, *Prunus*, *Ranunculus*, *Raphanus*, *Ribes*, *Ricinus*, *Rubus*, *Saccharum*, *Salpiglossis*, *Secale* (e.g., rye), *Senecio*, *Setaria*, *Sinapis*, *Solanum*, *Sorghum*, *Stenotaphrum*, *Theobroma*, *Trifolium*, *Trigonella*, *Triticum* (e.g., wheat), *Vicia*, *Vigna*, *Vitis*, *Zea* (e.g., corn), the *Olyreae*, the *Pharoideae* and many others.

For example, common crop plants which are targets of the present invention include corn, rice, triticale, rye, cotton, soybean, sorghum, wheat, oats, barley, millet, sunflower, canola, peas, beans, lentils, peanuts, yam beans, cowpeas, velvet beans, clover, alfalfa, lupine, vetch, lotus, sweet clover, wisteria, sweetpea and nut plants (e.g., walnut, pccan, etc).

In certain variations, naturally occurring in vivo recombination mechanisms of plants, agricultural microorganisms, or vector-host cells for intermediate replication can be used in conjunction with a collection of shuffled polynucleotide sequence variants having a desired phenotypic property to be optimized further; in this way, a natural recombination mechanism can be combined with intelligent selection of variants in an iterative manner to produce optimized

5 variants by "forced evolution", wherein the forced evolved variants are not expected  
to, nor are observed to, occur in nature, nor are predicted to occur at an appreciable  
frequency. The practitioner may further elect to supplement and/or the mutational  
10 drift by introducing intentionally mutated polynucleotide species suitable for  
5 shuffling, or portions thereof, into the pool of initial polynucleotide species and/or  
into the plurality of selected, shuffled polynucleotide species which are to be  
recombined. Mutational drift may also be supplemented by the use of mutagens (e.g.,  
15 chemical mutagens or mutagenic irradiation), or by employing replication conditions  
which enhance the mutation rate.

#### 10 Forced Evolution of Genes

The invention provides a means to evolve ADPGPP (S and/or L) gene  
20 variants and/or suitable host cells, as well as providing a model system for evaluating  
a library of agents to identify candidate agents that could find use as agricultural  
reagents for commercial applications. Such agents may exhibit selectivity for  
25 15 inhibition of a naturally occurring ADPGPP enzyme and may be substantially less  
effective at inhibiting a shuffled ADPGPP enzyme which has been evolved to be  
resistant to the agent.

#### ADPGPP Shuffling Combinations

30 Although the skilled artisan may select alternative shuffling strategies  
20 for enhancing ADPGPP enzyme properties, the following general combinations can  
be used:

35 I. Shuffling an ADPGPP from a first species of bacteria with an ADPGPP  
from a second species of bacteria. The resultant shufflants may be transformed into  
bacterial host cells which preferably lack endogenous ADPGPP activity (e.g., *E. coli*  
25 mutants glgC), algal cells, or plant cells for expression and selection. Phenotype  
40 selection of shufflants is typically performed by biochemical assay for ADPGPP, such  
as according to Preiss et al. (1966) Biochemistry 5: 1833; or other suitable assay  
method selected by the artisan, including microscopic detection of starch granules,  
specific gravity, iodine vapor colorimetry, or the like. Example bacteria for obtaining  
45 30 the ADPGPP gene(s) include *Rhodobacter sphaeroides*, *Rhodospirillum rubrum*,  
*Escherichia coli*, *Salmonella typhimurium*, and the like. A preferred host cell is a



5 strain of bacterium that is transformable and which lacks ADPGPP activity (e.g., glgC mutant of *E. coli*).

10 II. *Shuffling a parental bacterial ADPGPP encoding sequence with mutagenized variants thereof.* The resultant shufflants may be transformed into  
5 bacterial host cells which preferably lack endogenous ADPGPP activity (e.g., *E. coli*), algal cells, or plant cells for expression and selection. Phenotype selection of shufflants is typically performed by biochemical assay for ADPGPP activity or other  
15 suitable assay method selected by the artisan.

10 III. *Shuffling a L or S subunit from a first species of plant with a L subunit from a non-plant algae or bacterium, cyanobacteria.* The resultant shufflants may be transformed into host cells which preferably lack endogenous ADPGPP  
20 activity (e.g., *E. coli*), algal cells, or plant cells for expression and selection. Phenotype selection of shufflants is typically performed by biochemical assay for ADPGPP or other suitable assay method selected by the artisan. Example bacteria for  
25 the ADPGPP gene(s) include *Rhodobacter sphaeroides* (Falcone et al. (1998) *J. Bact.* 170: 5), *Rhodospirillum rubrum* (Falcone and Tabita (1993) *J. Bact.* 175: 5066; Falcone et al. (1991) *J. Bact.* 173: 2099), *Escherichia coli*, *Salmonella typhimurium*, and the like. Example cyanobacteria that can serve as a source of ADPGPP genes  
30 include *Synechococcus*, *Coccochloris peniocystis*, and *Aphanizomenon flos-aquae*. Example green algae that can serve as sources of ADPGPP genes include *Euglena gracilis*, *Chlamadomonas reinhardii*, and *Anacystis nidulans*. Example plants that can  
35 serve as sources for the L or S subunit genes include rice, maize, potato, wheat, rye, flax, cotton, pea, and the like.

25 IV. *Shuffling a plant L subunit from a first plant taxonomic species with a plant L subunit from a second plant taxonomic species.* The resultant shufflants may be transformed into host cells, which can preferably lack endogenous  
40 ADPGPP activity, but which fold and process higher plant ADPGPP subunits correctly for expression and selection, and generally encode and express a complementing plant S subunit, often encoded by a sequence derived from one or  
45 both of the higher plant species. Phenotype selection of shufflants is typically performed by iodine vapor visualization of blue-stained cells or by biochemical assay for ADPGPP or other suitable assay method selected by the artisan. Example higher  
50

plants that can serve as a source of ADPGPP L genes include, but are not limited to: *Zea mays* (C4), *Amaranthus hybridus* (C4), *Glycine max* (C3), and *Nicotiana tabacum* (C3).

V. *Shuffling a plant S subunit from a first plant taxonomic species with a plant S subunit from a second plant taxonomic species.* The resultant shufflants may be transformed into host cells, which can preferably lack endogenous ADPGPP activity, but which fold and process higher plant ADPGPP subunits correctly for expression and selection, and generally encode and express a complementing plant L subunit, often encoded by a sequence derived from one or both of the higher plant species. Phenotype selection of shufflants is typically performed by iodine vapor visualization of blue-stained cells or by biochemical assay for ADPGPP or other suitable assay method selected by the artisan. Example higher plants that can serve as a source of ADPGPP S genes include, but are not limited to: *Zea mays* (C4), *Amaranthus hybridus* (C4), *Glycine max* (C3), and *Nicotiana tabacum* (C3).

VI. *Shuffling a L or S subunit from a higher plant with mutagenized variants thereof.* An ADPGPP L or S gene ("parental gene") from a species of C3 or C4 plant is subjected to mutagenesis and shuffling/selection to generate a population of mutagenized shufflants which have substantial sequence identity to the parental gene. The population of mutagenized shufflants is transferred into a population of host cells wherein the mutagenized shufflants are expressed and the resultant transformed host cell population is selected or screened for an enhanced ADPGPP phenotype. Phenotype selection of shufflants is typically performed by biochemical assay for ADPGPP activity or other suitable assay method selected by the artisan.

#### Transcriptional Regulatory Sequences

Suitable transcriptional regulatory sequences include: cauliflower mosaic virus 19S and 35S promoters, NOS promoter, OCS promoter, rbcS promoter, *Brassica* heat shock promoter, synthetic promoters, non-plant promoters modified, if necessary, for function in plant cells, substantially any promoter that naturally occurs in a plant genome, promoters of plant viruses or Ti plasmids, tissue-preferential

5 promoters or cis-acting elements, light-responsive promoters or cis-acting elements  
(e.g., rbcS LRE), hormone-responsive cis-acting elements, developmental stage-  
specific promoters, organ specific promoters, cis-acting elements for promoters, viral  
10 promoters (e.g., from Tobacco Mosaic virus, Brome Mosaic Virus, Cauliflower  
Mosaic virus, and the like), and the like. In a variation, a transcriptional regulatory  
sequence from a first plant species is optimized for functionality in a second plant  
species by application of recursive sequence shuffling.

15 Transcriptional regulatory sequences for expression of shuffled  
ADPGPP sequences in chloroplasts is known in the art (Daniell et al. (1998) op.cit;  
10 O'Neill et al. (1993) The Plant Journal 3: 729; Maliga P (1993) op.cit), as are  
homologous recombination vectors.

#### 20 Host Cells for Screening ADPGPP Gene Shufflants

A variety of suitable host cells will be apparent to those skilled in the  
art. Of particular note, ADPGPP gene shufflants can be expressed in the glgC  
15 deletion mutant strain of *E. coli*, as well as higher taxonomic host cells. However,  
subunits from higher plants may not be processed correctly in bacterial host cells, so  
higher plant L and S gene shufflants may often be expressed for phenotype screening  
in plant cells, including mutant plant cell lines wherein an endogenous ADPGPP  
30 encoding gene has been functionally inactivated, preferably in homozygous format, to  
provide a plant cell substantially lacking endogenous ADPGPP activity, or the like.

#### 20 Transformation

35 The transformation of plants and protoplasts in accordance with the  
invention may be carried out in essentially any of the various ways known to those  
skilled in the art of plant molecular biology. See, in general, Methods in Enzymology  
25 Vol. 153 ("Recombinant DNA Part D") 1987, Wu and Grossman Eds., Academic  
Press, incorporated herein by reference. As used herein, the term transformation  
40 means alteration of the genotype of a host plant by the introduction of a nucleic acid  
sequence. The nucleic acid sequence need not necessarily originate from a different  
source, but it will, at some point, have been external to the cell into which it is to be  
45 introduced.

30 In one embodiment, the foreign nucleic acid is mechanically  
transferred by microinjection directly into plant cells by use of micropipettes.

5 Alternatively, the foreign nucleic acid may be transferred into the plant cell by using  
polyethylene glycol. This forms a precipitation complex with the genetic material  
that is taken up by the cell (e.g., by incubation of protoplasts with "naked DNA" in  
10 the presence of polyethyleneglycol)(Paszkowski et al., (1984) EMBO J. 3:2717-22;  
5 Baker et al (1985) Plant Genetics, 201-211; Li et al. (1990) Plant Molecular Biology  
Report 8(4)276-291].

15 In another embodiment of this invention, the introduced gene may be  
introduced into the plant cells by electroporation (Fromm et al., (1985) "Expression of  
Genes Transferred into Monocot and Dicot Plant Cells by Electroporation," Proc.  
10 Natl Acad. Sci. USA 82:5824, which is incorporated herein by reference). In this  
technique, plant protoplasts are electroporated in the presence of plasmids or nucleic  
20 acids containing the relevant genetic construct. Electrical impulses of high field  
strength reversibly permeabilize biomembranes allowing the introduction of the  
plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form a  
25 plant callus. Selection of the transformed plant cells with the transformed gene can  
be accomplished using phenotypic markers.

Cauliflower mosaic virus (CaMV) may also be used as a vector for  
introducing the foreign nucleic acid into plant cells (Hohn et al., (1982) "Molecular  
30 Biology of Plant Tumors," Academic Press, New York, pp.549-560; Howell, United  
20 States Patent No. 4,407,956). CaMV viral DNA genome is inserted into a parent  
bacterial plasmid creating a recombinant DNA molecule which can be propagated in  
bacteria. After cloning, the recombinant plasmid again may be cloned and further  
35 modified by introduction of the desired DNA sequence into the unique restriction site  
of the linker. The modified viral portion of the recombinant plasmid is then excised  
25 from the parent bacterial plasmid, and used to inoculate the plant cells or plants.

40 Another method of introduction of nucleic acid segments is high  
velocity ballistic penetration by small particles with the nucleic acid either within the  
matrix of small beads or particles, or on the surface (Klein et al., (1987) Nature  
45 327:70-73). Although typically only a single introduction of a new nucleic acid  
30 segment is required, this method particularly provides for multiple introductions.

A method of introducing the nucleic acid segments into plant cells is to  
infect a plant cell, an explant, a meristem or a seed with Agrobacterium tumefaciens

transformed with the segment. Under appropriate conditions known in the art, the transformed plant cells are grown to form shoots, roots, and develop further into plants. The nucleic acid segments can be introduced into appropriate plant cells, for example, by means of the Ti plasmid of Agrobacterium tumefaciens. The Ti plasmid is transmitted to plant cells upon infection by Agrobacterium tumefaciens, and is stably integrated into the plant genome (Horsch et al., (1984) "Inheritance of Functional Foreign Genes in Plants," Science, 233:496-498; Fraley et al., (1983) Proc. Natl. Acad. Sci. USA 80:4803).

Ti plasmids contain two regions essential for the production of transformed cells. One of these, named transfer DNA (T DNA), induces tumor formation. The other, termed virulent region, is essential for the introduction of the T DNA into plants. The transfer DNA region, which transfers to the plant genome, can be increased in size by the insertion of the foreign nucleic acid sequence without its transferring ability being affected. By removing the tumor-causing genes so that they no longer interfere, the modified Ti plasmid can then be used as a vector for the transfer of the gene constructs of the invention into an appropriate plant cell, such being a "disabled Ti vector."

All plant cells which can be transformed by Agrobacterium and whole plants regenerated from the transformed cells can also be transformed according to the invention so as to produce transformed whole plants which contain the transferred foreign nucleic acid sequence.

There are presently at least three different ways to transform plant cells with Agrobacterium: (1) co-cultivation of Agrobacterium with cultured isolated protoplasts; (2) transformation of cells or tissues with Agrobacterium, or (3) transformation of seeds, apices or meristems with Agrobacterium.

Method (1) uses an established culture system that allows culturing protoplasts and plant regeneration from cultured protoplasts.

Method (2) implies (a) that the plant cells or tissues can be transformed by Agrobacterium and (b) that the transformed cells or tissues can be induced to regenerate into whole plants.

Method (3) uses micropropagation. In the binary system, to have infection, two plasmids are needed: a T-DNA containing plasmid and a vir plasmid.

Any one of a number of T-DNA containing plasmids can be used, the main issue being that one be able to select independently for each of the two plasmids.

After transformation of the plant cell or plant, those plant cells or plants transformed by the Ti plasmid so that the desired DNA segment is integrated can be selected by an appropriate phenotypic marker. These phenotypic markers include, but are not limited to, antibiotic resistance, herbicide resistance or visual observation. Other phenotypic markers are known in the art and may be used in this invention.

#### Protoplast Transformation

Numerous protocols for establishment of transformable protoplasts from a variety of plant types and subsequent transformation of the cultured protoplasts are available in the art and are incorporated herein by general reference. For examples, see Hashimoto et al. (1990) Plant Physiol. 93: 857; Plant Protoplasts, Fowke LC and Constabel F, eds., CRC Press (1994); Saunders et al. (1993) Applications of Plant In Vitro Technology Symposium, UPM, 16-18 Nov. 1993; and Lyznik et al. (1991) BioTechniques 10: 295, each of which is incorporated herein by reference).

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred foreign gene. Some suitable plants include, for example, species from the genera Fragaria, Lotus, Medicago, Onobrychis, Trifolium, Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum, Hyoscyamus, Lycopersicon, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Cichorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Hererocallis, Nemesia, Pelargonium, Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Lolium, Zea, Triticum, Sorghum, and Datura.

It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major cereal crop species, sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Limited knowledge presently exists on whether all of these plants can be transformed by Agrobacterium. Species which are a natural plant host for Agrobacterium may be transformable in

5 vitro. Although monocotyledonous plants, and in particular, cereals and grasses, are  
not natural hosts to Agrobacterium, work to transform them using Agrobacterium has  
also been successfully carried out by numerous investigators (Honykas-Van Slogteren  
et al., (1984) Nature 311:763-764; Hernalsteens et al., (1984) EMBO J. 3:3039-41;  
10 5 Byteiber, et al. (1987) Proc. Natl. Acad. Sci. USA: 5345-5349; Graves and Goldman,  
(1986) Plant Mol. Biol 7: 43-50; Grimsley et al. (1988) Biochemistry 6: 185-189; WO  
86/03776; Shimamoto et al. Nature (1989) 338: 274-276). Monocots may also be  
transformed by techniques or with vectors other than Agrobacterium. For example,  
15 monocots have been transformed by electroporation (Fromm et al. [1986] Nature  
10 319:791-793; Rhoads et al. Science [1988] 240: 204-207), direct gene transfer (Baker  
et al. [1985] Plant Genetics 201-211), by using pollen-mediated vectors (EP 0 270  
20 356), and by injection of DNA into floral tillers (de la Pena et al. [1987], Nature  
325:274-276). Additional plant genera that may be transformed by Agrobacterium  
include Chrysanthemum, Dianthus, Gerbera, Euphorbia, Pelargonium, Ipomoea,  
15 Passiflora, Cyclamen, Malus, Prunus, Rosa, Rubus, Populus, Santalum, Allium,  
25 Lilium, Narcissus, Ananas, Arachis, Phaseolus and Pisum.

#### Chloroplast Transformation

As the ADPGPP enzyme of higher plants is encoded in the nuclear  
30 genome and expressed with a fused chloroplast transit sequence peptide (CTS) to  
20 facilitate translocation of the ADPGPP subunits into chloroplasts, it can be  
advantageous to transform the shufflant ADPGPP encoding sequences into  
chloroplasts if the host cells are derived from higher plants. Numerous methods are  
35 available in the art to accomplish the chloroplast transformation and expression  
(Daniell et al. (1998) op.cit; O'Neill et al. (1993) The Plant Journal 3: 729; Maliga P  
25 (1993) op.cit). The expression construct comprises a transcriptional regulatory  
sequence functional in plants operably linked to a polynucleotide encoding an  
40 enhanced ADPGPP protein subunit. With respect to polynucleotide sequences  
encoding ADPGPP subunit proteins, it may be desirable to express such encoding  
sequences in plastids, such as chloroplasts, for appropriate transcription, translation,  
45 and processing. With reference to expression cassettes which are designed to function  
30 in chloroplasts, such as an expression cassette encoding a subunit of ADPGPP in a  
higher plant, the expression cassette comprises the sequences necessary to ensure

5 expression in chloroplasts - typically the subunit encoding sequence is flanked by two  
regions of homology to the plastid genome so as to effect a homologous  
recombination with the chloroplastid genome; often a selectable marker gene is also  
present within the flanking plastid DNA sequences to facilitate selection of  
10 5 genetically stable transformed chloroplasts in the resultant transplastonic plant cells  
(see Maliga P (1993) TIBTECH 11: 101; Daniell et al. (1998) Nature Biotechnology  
16: 346, and references cited therein).

#### 15 Recovery of Selected Polynucleotide Sequences

A variety of selection and screening methods will be apparent to those  
10 skilled in the art, and will depend upon the particular phenotypic properties that are  
desired. The selected shuffled genetic sequences can be recovered for further  
shuffling or for direct use by any applicable method, including but not limited to:  
20 recovery of DNA, RNA, or cDNA from cells (or PCR-amplified copies thereof) from  
cells or medium, recovery of sequences from host chromosomal DNA or PCR-  
15 amplified copies thereof, recovery of episome (e.g., expression vector) such as a  
plasmid, cosmid, viral vector, artificial chromosome, and the like, or other suitable  
25 recovery method known in the art.

Any suitable art-known method, including RT-PCR or PCR, can be  
30 used to obtain the selected shufflant sequence(s) for subsequent manipulation and  
20 shuffling.

#### Backcrossing

After a desired ADPGPP phenotype is acquired to a satisfactory extent  
35 by a selected shuffled gene or portion thereof, it is often desirable to remove  
mutations which are not essential or substantially important to retention of the desired  
25 phenotype ("superfluous mutations"). This is particularly desirable when the shuffled  
gene sequence is to be reintroduced back into a higher plant, as it is often preferred to  
40 harmonize the shufflant ADPGPP subunit sequence with the endogenous ADPGPP  
subunit sequence in the higher plant taxonomic species genome while retaining the  
desired ADPGPP phenotype obtained from the iterative shuffling/selection process.  
45 30 Superfluous mutations can be removed by backcrossing, which is shuffling the  
selected shuffled ADPGPP L gene(s) with one or more parental ADPGPP L gene  
and/or naturally-occurring ADPGPP L gene(s) (or portions thereof) and selecting the



5 resultant collection of shufflants for those species that retain the desired phenotype.  
The same process may be employed for the ADPGPP S genes. By employing this  
method, typically in two or more recursive cycles of shuffling against parental or  
naturally-occurring viral genome(s) (or portions thereof) and selection for retention of  
10 the desired ADPGPP phenotype, it is possible to generate and isolate selected  
shufflants which incorporate substantially only those mutations necessary to confer  
the desired phenotype, whilst having the remainder of the genome (or portion thereof)  
consist of sequence which is substantially identical to the parental (or wild-type)  
15 sequence(s). As one example of backcrossing, a potato ADPGPP subunit gene (small  
or large subunit) can be shuffled and selected for the capacity to substantially function  
in any Angiosperm plant cells; the resultant selected shufflants can be backcrossed  
with one or more ADPGPP genes of a particular plant species and selected for the  
20 capacity to retain the capacity to confer the phenotype. After several cycles of such  
backcrossing, the backcrossing will yield gene(s) which contain the mutations  
necessary for the desired phenotype, and will otherwise have a genomic sequence  
substantially identical to the genome(s) of the host genome.

Isolated components (e.g., genes, regulatory sequences, replication  
origins, and the like) can be optimized and then backcrossed with parental sequences  
so as to obtain optimized components which are substantially free of superfluous  
25 mutations.

#### Transgenic Hosts

Transgenes and expression vectors to express shufflant  
ADPGPP sequences can be constructed by any suitable method known in the art; by  
either PCR or RT-PCR amplification from a suitable cell type or by ligating or  
25 amplifying a set of overlapping synthetic oligonucleotides; publicly available  
sequence databases and the literature can be used to select the polynucleotide  
sequence(s) to encode the specific protein desired, including any mutations,  
consensus sequence, or mutation kernel desired by the practitioner. The coding  
sequence(s) are operably linked to a transcriptional regulatory sequence and, if  
45 desired, an origin of replication. Antisense or sense-suppression transgenes and  
genetic sequences can be optimized or adapted for particular host cells and organisms  
by the described methods.

The transgene(s) and/or expression vectors are transferred into host cells, protoplasts, pluripotent embryonic plant cells, microbes, or fungi by a suitable method, such as for example lipofection, electroporation, microinjection, biolistics, *Agrobacterium tumefaciens* transduction of Ti plasmid, calcium phosphate precipitation, PEG-mediated DNA uptake, electroporation, electrofusion, or other method. Stable transfectant host cells can be prepared by art-known methods, as can transgenic cell lines.

#### Target Plants

As used herein, "plant" refers to either a whole plant, a plant part, a plant cell, or a group of plant cells. The class of plants which can be used in the method of the invention is generally as broad as the class of higher plants amenable to protoplast transformation techniques, including both monocotyledonous and dicotyledonous plants. It includes plants of a variety of ploidy levels, including polyploid, diploid and haploid, and may employ non-regenerable cells for certain aspects which do not require development of an adult plant for selection or in vivo shuffling.

Preferred plants for the transformation and expression shuffled genes of this invention include agronomically and horticulturally important species. Such species include, but are not restricted to members of the families: *Graminae* (including corn, rye, triticale, barley, millet, rice, wheat, oats, etc.); *Leguminosae* (including pea, beans, lentil, peanut, yam bean, cowpeas, velvet beans, soybean, clover, alfalfa, lupine, vetch, lotus, sweet clover, wisteria, and sweetpea); *Compositae* (the largest family of vascular plants, including at least 1,000 genera, including important commercial crops such as sunflower) and *Rosaciae* (including raspberry, apricot, almond, peach, rose, etc.), as well as nut plants (including, walnut, pecan, hazelnut, etc.)

Targets for the invention include plants from the genera: *Agrostis*, *Allium*, *Antirrhinum*, *Apium*, *Arachis*, *Asparagus*, *Atropa*, *Avena* (e.g., oats), *Bambusa*, *Brassica*, *Bromus*, *Browaalia*, *Camellia*, *Cannabis*, *Capsicum*, *Cicer*, *Chenopodium*, *Chichorium*, *Citrus*, *Coffea*, *Coix*, *Cucumis*, *Curcubita*, *Cynodon*, *Dactylis*, *Datura*, *Daucus*, *Digitalis*, *Dioscorea*, *Elaeis*, *Eleusine*, *Festuca*, *Fragaria*, *Geranium*, *Glycine*, *Helianthus*, *Heterocallis*, *Hevea*, *Hordeum* (e.g., barley), *Hyoscyamus*, *Ipomoeu*, *Lactuca*, *Lens*, *Lilium*, *Linum*, *Lolium*, *Lotus*, *Lycopersicon*,

*Majorana, Malus, Mangifera, Manihot, Medicago, Nemesia, Nicotiana, Onobrychis, Oryza* (e.g., rice), *Panicum, Pelargonium, Pennisetum* (e.g., millet), *Petunia, Pisum, Phaseolus, Phleum, Poa, Prunus, Ranunculus, Raphanus, Ribes, Ricinus, Rubus, Saccharum, Salpiglossis, Secale* (e.g., rye), *Senecio, Setaria, Sinapis, Solanum, Sorghum, Stenotaphrum, Theobroma, Trifolium, Trigonella, Triticum* (e.g., wheat), *Vicia, Vigna, Vitis, Zea* (e.g., corn), and the *Olyreae*, the *Pharoideae* and many others.

Common crop plants which are targets of the present invention include corn, rice, tritcale, rye, cotton, soybean, sorghum, wheat, oats, barley, millet, sunflower, canola, peas, beans, lentils, peanuts, yam beans, cowpeas, velvet beans, clover, alfalfa, lupine, vetch, lotus, sweet clover, wisteria, sweetpea and nut plants (e.g., walnut, pecan, etc).

#### Regeneration

Normally, regeneration will be involved in obtaining a whole plant from the transformation process. The term "transgenote" refers to the immediate product of the transformation process and to resultant whole transgenic plants.

The term "regeneration" as used herein, means growing a whole plant from a plant cell, a group of plant cells, a plant part or a plant piece (e.g. from a protoplast, callus, or tissue part).

Plant regeneration from cultural protoplasts is described in Evans et al., "Protoplasts Isolation and Culture," Handbook of Plant Cell Cultures 1:124-176 (MacMillan Publishing Co. New York 1983); M.R. Davey, "Recent Developments in the Culture and Regeneration of Plant Protoplasts," Protoplasts, (1983) - Lecture Proceedings, pp.12-29, (Birkhauser, Basel 1983); P.J. Dale, "Protoplast Culture and Plant Regeneration of Cereals and Other Recalcitrant Crops," Protoplasts (1983) - Lecture Proceedings, pp. 31-41, (Birkhauser, Basel 1983); and H. Binding, "Regeneration of Plants," Plant Protoplasts, pp.21-73, (CRC Press, Boca Raton 1985).

Additional details regarding plant regeneration are found in Jones (ed) (1995) Plant Gene Transfer and Expression Protocols-- Methods in Molecular Biology, Volume 49 Humana Press Towata NJ; Payne et al. (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New York, NY (Payne); Gamborg and Phillips (eds) (1995) Plant Cell, Tissue and Organ Culture;

5                    Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg  
New York) (Gamborg) and in Croy, (ed.) (1993) Plant Molecular Biology.

10                    5                    Regeneration from protoplasts varies from species to species of plants,  
but generally a suspension of transformed protoplasts containing copies of the  
exogenous sequence is first made. In certain species embryo formation can then be  
15                    10                    induced from the protoplast suspension, to the stage of ripening and germination as  
natural embryos. The culture media will generally contain various amino acids and  
hormones, such as auxin and cytokinins. It is sometimes advantageous to add  
glutamic acid and proline to the medium, especially for such species as corn and  
alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration  
20                    will depend on the medium, on the genotype, and on the history of the culture. If  
these three variables are controlled, then regeneration is fully reproducible and  
repeatable.

15                    15                    Regeneration also occurs from plant callus, explants, organs or parts.  
Transformation can be performed in the context of organ or plant part regeneration.  
25                    See, Methods in Enzymology, *supra*; also Methods in Enzymology, Vol. 118; and  
Klee et al., (1987) Annual Review of Plant Physiology, 38:467-486.

30                    20                    In vegetatively propagated crops, the mature transgenic plants are  
propagated by the taking of cuttings or by tissue culture techniques to produce  
multiple identical plants for trialling, such as testing for production characteristics.  
Selection of desirable transgenes is made and new varieties are obtained thereby,  
and propagated vegetatively for commercial sale.

35                    25                    In seed propagated crops, the mature transgenic plants are self crossed  
to produce a homozygous inbred plant. The inbred plant produces seed containing  
the gene for the newly introduced foreign gene activity level. These seeds can be  
40                    grown to produce plants that would produce the selected phenotype.

45                    30                    The inbreds according to this invention can be used to develop new  
hybrids. In this method a selected inbred line is crossed with another inbred line to  
produce the hybrid. The offspring resulting from the first experimental crossing of  
two parents is known in the art as the F1 hybrid, or first filial generation. Of the two  
parents crossed to produce F1 progeny according to the present invention, one or both  
50                    parents can be transgenic plants.

5 Parts obtained from the regenerated plant, such as flowers, seeds,  
leaves, branches, fruit, and the like are covered by the invention, provided that these  
parts comprise cells which have been so transformed. Progeny and variants, and  
mutants of the regenerated plants are also included within the scope of this invention,  
10 5 provided that these parts comprise the introduced DNA sequences. Progeny and  
variants, and mutants of the regenerated plants are also included within the scope of  
this invention.

15 The following example is given to illustrate the invention, but are not  
to be limiting thereof.

10 EXPERIMENTAL EXAMPLE

EXAMPLE 1: Shuffling ADP-glucose pyrophosphorylase

20 Genes coding for ADP-glucose pyrophosphorylase (ADPGPP) from E.  
coli are isolated using primers designed from published sequence in the Genbank. A  
genomic DNA library of E. coli is used as a source for the ADPGPP gene. Similarly,  
15 ADPGPP genes from other microorganisms are isolated including from  
25 cyanobacteria. All of these prokaryotes have a single subunit ADPGPP (Preiss J,  
(1996) Biotechnology Annual Review Vol. 2, pp259-279).

The ADPGPP genes from various microorganisms, which have at least  
30 70 percent nucleotide sequence identity are shuffled according to published  
20 procedures. Briefly, this procedure involves random fragmentation of the genes with  
DNase I and selecting nucleotide fragments of 100-300 bp. The fragments are  
reassembled based on sequence similarity by primerless PCR. Recombination as well  
35 as variable levels of mutations that are introduced by the PCR reaction generate the  
diversity. The assembled genes is cloned into a starch minus  
25 E. coli mutant that lacks ADPGPP such as LCB618 (available at the Coli Genetics  
Stock Center at Yale). Transformed colonies expressing a functional ADPGPP are  
40 screened for production of glycogen by iodine staining (Greene TW et al. (1996)  
PNAS 93: 1509-1513). Those colonies staining dark blue (greater starch content) are  
presumed to contain deregulated ADPGPP. Colonies expressing shuffled ADPGPP  
45 30 genes are selected and grown in larger amounts in liquid culture and assayed for  
specific properties (Meyer et al. (1998) Archives Biochem. Biophys. Pp152-159)  
relative to the wildtype enzyme, such as: (a) insensitivity to activation by fructose-1,

5 6-bisphosphate (FBP) (b) desensitized to inhibition by AMP and inorganic phosphate  
(c) decreased  $K_m$  for the two substrates, glucose-1-phosphate and ATP (d) increased  
10  $V_{max}$ . Genes from those clones expressing one or more of the desired properties  
5 mentioned above are iteratively shuffled in order to achieve optimization of one or  
more of the properties mentioned above. The optimized gene, after appropriate  
modification, is used to transform the desired crop species in order to deregulate and  
increase starch biosynthesis in various tissues including tubers and seeds.

15 Plant genes coding for ADPGPP are cloned into *E. coli* (Iglesias A et  
al. *J. Biol Chem* 268: 1081-1086) and shuffled as described above, to optimize the  
10 desired properties. The plant enzyme is composed of two subunits, the small catalytic  
and the large regulatory subunit. Both genes are shuffled individually or in  
20 combination. Selection is done in *E. coli* as described above. Enzyme assays can be  
performed for analysis of properties as described in literature (Meyer et al. (1998)  
*Archives Biochem. Biophys.* Pp152-159). A difference between the plant and  
15 bacterial enzyme is that the activator is 3-phosphoglycerate and the inhibitor is  
25 inorganic phosphate.

#### Integrated Systems

The present invention provides computers, computer readable media  
30 and integrated systems comprising character strings corresponding to shuffled  
20 ADPGPP enzyme and corresponding enzyme-encoding nucleic acids. These  
sequences can be manipulated by in silico shuffling methods, or by standard sequence  
alignment or word processing software.

35 For example, different types of similarity and considerations of various  
stringency and character string length can be detected and recognized in the  
25 integrated systems herein. For example, many homology determination methods have  
been designed for comparative analysis of sequences of biopolymers, for spell-  
40 checking in word processing, and for data retrieval from various databases. With an  
understanding of double-helix pair-wise complement interactions among 4 principal  
nucleobases in natural polynucleotides, models that simulate annealing of  
45 complementary homologous polynucleotide strings can also be used as a foundation  
30 of sequence alignment or other operations typically performed on the character  
strings corresponding to the sequences herein (e.g., word-processing manipulations,

5 construction of figures comprising sequence or subsequence character strings, output  
tables, etc.). An example of a software package with algorithms for calculating  
sequence similarity is BLAST, which can be adapted to the present invention by  
inputting character strings corresponding to the sequences herein.

10 5 BLAST is described in Altschul *et al.*, *J. Mol. Biol.* 215:403-410  
(1990). Software for performing BLAST analyses is publicly available through the  
National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This  
15 algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying  
short words of length W in the query sequence, which either match or satisfy some  
positive-valued threshold score T when aligned with a word of the same length in a  
10 database sequence. T is referred to as the neighborhood word score threshold  
(Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for  
initiating searches to find longer HSPs containing them. The word hits are then  
extended in both directions along each sequence for as far as the cumulative  
20 alignment score can be increased. Cumulative scores are calculated using, for  
nucleotide sequences, the parameters M (reward score for a pair of matching residues;  
always > 0) and N (penalty score for mismatching residues; always < 0). For amino  
25 acid sequences, a scoring matrix is used to calculate the cumulative score. Extension  
of the word hits in each direction are halted when: the cumulative alignment score  
falls off by the quantity X from its maximum achieved value; the cumulative score  
goes to zero or below, due to the accumulation of one or more negative-scoring  
30 residue alignments; or the end of either sequence is reached. The BLAST algorithm  
parameters W, T, and X determine the sensitivity and speed of the alignment. The  
BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of  
25 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both  
strands. For amino acid sequences, the BLASTP program uses as defaults a  
wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix  
(see Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

45 30 An additional example of a useful sequence alignment algorithm is  
PILEUP. PILEUP creates a multiple sequence alignment from a group of related  
sequences using progressive, pairwise alignments. It can also plot a tree showing the  
clustering relationships used to create the alignment. PILEUP uses a simplification of

5 the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360  
(1987). The method used is similar to the method described by Higgins & Sharp,  
10 *CABIOS* 5:151-153 (1989). The program can align, e.g., up to 300 sequences of a  
maximum length of 5,000 letters. The multiple alignment procedure begins with the  
15 pairwise alignment of the two most similar sequences, producing a cluster of two  
aligned sequences. This cluster can then be aligned to the next most related sequence  
or cluster of aligned sequences. Two clusters of sequences can be aligned by a simple  
extension of the pairwise alignment of two individual sequences. The final alignment  
20 is achieved by a series of progressive, pairwise alignments. The program can also be  
used to plot a dendrogram or tree representation of clustering relationships. The  
program is run by designating specific sequences and their amino acid or nucleotide  
coordinates for regions of sequence comparison.

The shuffled enzymes of the invention, or corresponding coding  
nucleic acids, are optionally sequenced and the sequences aligned to provide structure-  
25 function information. For example, the alignment of shuffled sequences which are  
selected for conversion activity against the same target provides an indication of  
which residues are relevant for conversion of the target (i.e., conserved residues are  
likely more important for activity than non-conserved residues).

Standard desktop applications such as word processing software (e.g.,  
30 Microsoft Word™ or Corel WordPerfect™) and database software (e.g., spreadsheet  
software such as Microsoft Excel™, Corel Quattro Pro™, or database programs such  
as Microsoft Access™ or Paradox™) can be adapted to the present invention by  
35 inputting character strings corresponding to shuffled ADPGPP enzymes (or  
corresponding coding nucleic acids), e.g., shuffled by the methods herein. For  
example, the integrated systems can include the foregoing software having the  
40 appropriate character string information, e.g., used in conjunction with a user  
interface (e.g., a GUI in a standard operating system such as a Windows, Macintosh  
or LINUX system) to manipulate strings of characters. As noted, specialized  
alignment programs such as BLAST or PILEUP can also be incorporated into the  
45 systems of the invention for alignment of nucleic acids or proteins (or corresponding  
character strings).

Integrated systems for analysis in the present invention typically



5

10

15

20

25

30

35

40

45

50

55

5

10

15

20

25

30

include a digital computer with software for aligning or manipulating sequences, as well as data sets entered into the software system comprising any of the sequences herein. The computer can be, *e.g.*, a PC (Intel x86 or Pentium chip- compatible DOS<sup>TM</sup>, OS2<sup>TM</sup> WINDOWST<sup>TM</sup> WINDOWS NT<sup>TM</sup>, WINDOWS95<sup>TM</sup>, WINDOWS98<sup>TM</sup> LINUX based machine, a MACINTOSH<sup>TM</sup>, Power PC, or a UNIX based (*e.g.*, SUN<sup>TM</sup> work station) machine) or other commercially common computer which is known to one of skill. Software for aligning or otherwise manipulating sequences is available, or can easily be constructed by one of skill using a standard programming language such as Visual basic, Fortran, Basic, Java, or the like.

Any controller or computer optionally includes a monitor which is often a cathode ray tube ("CRT") display, a flat panel display (*e.g.*, active matrix liquid crystal display, liquid crystal display), or others. Computer circuitry is often placed in a box which includes numerous integrated circuit chips, such as a microprocessor, memory, interface circuits, and others. The box also optionally includes a hard disk drive, a floppy disk drive, a high capacity removable drive such as a writeable CD-ROM, and other common peripheral elements. Inputting devices such as a keyboard or mouse optionally provide for input from a user and for user selection of sequences to be compared or otherwise manipulated in the relevant computer system.

The computer typically includes appropriate software for receiving user instructions, either in the form of user input into a set parameter fields, *e.g.*, in a GUI, or in the form of preprogrammed instructions, *e.g.*, preprogrammed for a variety of different specific operations. The software then converts these instructions to appropriate language for instructing the system to carry out any desired operation.

In one aspect, the computer system is used to perform "in silico" shuffling of character strings. A variety of such methods are set forth in "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov and Stemmer, filed February 5, 1999 (USSN 60/118854) and "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov and Stemmer, filed October 12, 1999 (USSN 09/416,375). In brief, in the context of the present

invention, genetic operators are used in genetic algorithms as described in the '375 application to change given ADPGPP sequences, e.g., by mimicking genetic events such as mutation, recombination, death and the like. Multi-dimensional analysis to optimize sequences can be also be performed in the computer system, e.g., as described in the '375 application.

A digital system can also instruct an oligonucleotide synthesizer to synthesize oligonucleotides, e.g., used for ADPGPP gene reconstruction or recombination, or to order oligonucleotides from commercial sources (e.g., by printing appropriate order forms or by linking to an order form on the internet).

The digital system can also include output elements for controlling nucleic acid synthesis (e.g., based upon a sequence or an alignment of a shuffled enzyme as herein), i.e., an integrated system of the invention optionally includes an oligonucleotide synthesizer or an oligonucleotide synthesis controller. The system can include other operations which occur downstream from an alignment or other operation performed using a character string corresponding to a sequence herein, e.g., as noted above with reference to assays.

#### Combination Shuffling

One aspect of the present invention is the combinatorial shuffling of ADGPP with enzymes that affect carbon fixation. For example, one aspect of the present invention involves separately or simultaneously shuffling ADGPP in combination with carbon fixation enzymes such as ribulose 1,5-bisphosphate carboxylase/oxygenase ("Rubisco"; EC 4.1.1.39), or with any Calvin cycle enzyme or Krebs cycle enzyme. Considerable detail regarding Rubisco and Calvin and Krebs cycle enzymes and shuffling of such enzymes to improve carbon fixation is found in commonly assigned U.S. Patent Application U.S.S.N. 60/107,756 and 60/153,093 entitled "MODIFIED RIBULOSE BISPHOSPHATE CARBOXYLASE/OXYGENASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES," filed on 10 November 1998 and September 9, 1999, respectively and in "MODIFIED RIBULOSE BISPHOSPHATE CARBOXYLASE/OXYGENASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES," by Stemmer et al., co-filed November 9, 1999 (Attorney Docket number 02-292-2US/PC). Shuffled ADPGPP genes and shuffled Rubisco genes are

optionally co-expressed in a cell or organism such as a plant to increase starch production and carbon fixation.

Similarly, shuffled ADPGPP genes can be expressed with shuffled Phosphoenolpyruvate (PEP) carboxylase (PEPC; EC 4.1.1.31) genes to improve carbon fixation and starch production. Considerable detail regarding PEPC gene shuffling is found in commonly assigned U.S. Patent Application U.S.S.N. 60/107,757 entitled "MODIFIED PHOSPHOENOLPYRUVATE CARBOXYLASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES" filed on 10 November 1998 (Attorney Docket Number 018097-029100US) and in "MODIFIED PHOSPHOENOLPYRUVATE CARBOXYLASE FOR IMPROVEMENT AND OPTIMIZATION OF PLANT PHENOTYPES" co-filed on 9 November 1999 (Attorney Docket Number 02-0291-1US/PC) by Stemmer and Subramanian. Shuffled ADGPP genes and shuffled PEPC genes are optionally co-expressed in a cell or organism such as a plant to increase starch production and carbon fixation. Of course, shuffled Rubisco, ADPGPP, and PEPC can all be expressed together in a cell or organism such as a plant to increase carbon fixation, starch production, and the like.

5

In a further aspect, the present invention provides for the use of any apparatus, apparatus component, composition or kit herein, for the practice of any method or assay herein, and/or for the use of any apparatus or kit to practice any assay or method herein.

10

5

15

The foregoing description of the preferred embodiments of the present invention has been presented for purposes of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise form disclosed, and many modifications and variations are possible in light of the above teaching.

10

Such modifications and variations which may be apparent to a person skilled in the art are intended to be within the scope of this invention.

20

*All publications and patent applications herein are incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.*

25

30

35

40

45

50

55

Claims

5

10

15

20

25

30

35

40

45

50

55

WHAT IS CLAIMED IS:

1. A method for obtaining an isolated polynucleotide encoding an enhanced ADPGPP protein having ADPGPP catalytic activity wherein the ADPGPP enzymatic phenotype is significantly different than a protein encoded by a parental polynucleotide encoding a naturally-occurring ADPGPP enzyme, the method comprising:

recombining sequences of a plurality of parental polynucleotide species encoding at least one ADPGPP sequence under conditions suitable for sequence shuffling to form a resultant library of sequence-shuffled ADPGPP polynucleotides;

transferring said library into a plurality of host cells forming a library of transformants wherein sequence-shuffled ADPGPP polynucleotides are expressed;

assaying individual or pooled transformants for ADPGPP catalytic activity to determine the relative or absolute ADPGPP enzymatic phenotype and isolating a transformant having a ADPGPP enzymatic phenotype significantly different than parental ADPGPP, thereby identifying at least one enhanced transformant that expresses a ADPGPP enzyme activity which has a significantly altered compared to the ADPGPP activity encoded by the parental sequence(s);

recovering the sequence-shuffled ADPGPP polynucleotide from at least one enhanced transformant.

2. The method of claim 1, further comprising the step of subjecting a recovered sequence-shuffled ADPGPP polynucleotide encoding an enhanced ADPGPP to at least one subsequent round of recursive shuffling and selection, wherein said recovered sequence-shuffled ADPGPP polynucleotide is used as at least one parental sequence for subsequent shuffling.

3. The method of claim 1, wherein selection comprises assaying individual or pooled transformants for ADPGPP catalytic activity to determine the relative or absolute  $K_m$  for substrate and identifying at least one enhanced transformant that expresses an ADPGPP activity which has a significantly lower  $K_m$  for substrate than the ADPGPP activity encoded by the parental sequence(s).

5                   4.     The method of claim 1, wherein selection comprises assaying  
individual or pooled transformants for ADPGPP catalytic activity to determine the  
relative or absolute  $K_m$  for inhibitor thereby identifying at least one enhanced  
transformant that expresses an ADPGPP activity which has a significantly higher  $K_m$   
10                   5     for inhibitor than the ADPGPP activity encoded by the parental sequence(s).

15                   5.     The method of claim 1, wherein selection comprises assaying  
individual or pooled transformants for ADPGPP catalytic activity to determine the  
relative or absolute  $K_m$  for activator thereby identifying at least one enhanced  
10                   10    transformant that expresses an ADPGPP activity which has a significantly lower  $K_m$   
for activator than the ADPGPP activity encoded by the parental sequence(s).

20                   6.     The method of claim 1, wherein selection comprises assaying samples  
of individual transformants and their clonal progeny which are isolated into discrete  
15                   15    reaction vessels for ADPGPP activity assay, or are assayed in situ.

25                   7.     The method of claim 1, wherein the step of recombining the sequences of  
the plurality of parental polynucleotide species is performed in vitro, in vivo or in  
30                   20    silico.

35                   8.     The method of claim 1, wherein the host cell comprises a non-  
photosynthetic bacterium lacking an endogenous ADPGPP activity and is  
transformed with an expression cassette encoding a shufflant ADPGPP protein,  
optionally including an expression cassette encoding a complementing ADPGPP  
25                   25    subunit and, wherein selection comprises culturing the population of transformed host  
cells in the presence of iodine for a suitable incubation period, determining the  
40                   40    relative or absolute amount of iodine-reacted starch in each transformed host cell and  
its clonal progeny relative to the amount of iodine-reacted starch in untransformed  
host cells cultured under equivalent conditions.

45                   30                   9.     The method of claim 8, wherein the host cells harbor expression  
cassettes encoding a complementing L subunit and the library comprises shuffled S  
50

subunit encoding sequences.

10. A plant cell protoplast and clonal progeny thereof containing a sequence-shuffled polynucleotide encoding a ADPGPP subunit which is not encoded by the naturally occurring genome of the plant cell protoplast.

11. A collection of plant cell protoplasts transformed with a library of sequence-shuffled ADPGPP subunit polynucleotides in expressible form.

12. A regenerated plant containing at least one species of replicable or integrated polynucleotide comprising a sequence-shuffled portion and encoding an ADPGPP subunit polypeptide.

13. A regenerated plant containing a polynucleotide expression cassette encoding a shuffled ADPGPP gene.

14. A regenerated plant of claim 13, further comprising a polynucleotide expression cassette encoding a shuffled bacterial or algal ADPGPP gene.

15. A polynucleotide comprising: (1) a sequence encoding a shuffled ADPGPP subunit gene linked to (2) a selectable marker gene which affords a means of selection when expressed in chloroplasts, and, optionally, flanked by (3) an upstream flanking recombinogenic sequence having sufficient sequence identity to a chloroplast genome sequence to mediate efficient recombination and (4) a downstream flanking recombinogenic sequence having sufficient sequence identity to a chloroplast genome sequence to mediate efficient recombination.

16. A polynucleotide of claim 14, wherein the polynucleotide encodes an enhanced ADPGPP protein having ADPGPP catalytic activity wherein the  $K_m$  for substrate is significantly lower than a protein encoded by a parental polynucleotide encoding a naturally-occurring plant ADPGPP enzyme.



5 17. A polynucleotide of claim 14, wherein the polynucleotide encodes an enhanced ADPGPP protein having ADPGPP catalytic activity wherein the  $K_m$  for inhibitor is significantly higher than a protein encoded by a parental polynucleotide encoding a naturally-occurring ADPGPP enzyme or subunit.

10 5 18. A polynucleotide of claim 14, wherein the polynucleotide encodes an enhanced ADPGPP protein having ADPGPP catalytic activity wherein: (1) the  $K_m$  for substrate is significantly lower than a protein encoded by a parental polynucleotide encoding a naturally-occurring ADPGPP enzyme, (2) the  $K_m$  for inhibitor is significantly higher than a protein encoded by a parental polynucleotide encoding a naturally-occurring ADPGPP enzyme, and/or (3) the  $K_m$  for activator is significantly lower than a protein encoded by a parental polynucleotide encoding a naturally-occurring ADPGPP enzyme.

20 19. A method of producing a recombinant cell having an elevated starch production activity, the method comprising:

25 15 (A) recombining one or more first ADPGPP enzyme coding nucleic acid, or a homologue thereof, with one or more homologous first nucleic acid to produce a library of recombinant first enzyme nucleic acid homologues;

30 20 (B) optionally repeating step (A) one or more times using one or more members of the library of recombinant first enzyme nucleic acid homologues as the one or more first ADPGPP enzyme coding nucleic acid, or the homologue thereof, or as the one or more first homologous nucleic acid, thereby producing a diversified library of recombinant first enzyme nucleic acid homologues;

35 25 (C) selecting the library of recombinant first enzyme nucleic acid homologues or the diversified library of recombinant first enzyme nucleic acid homologues for one or more of: an increased or decreased ADPGPP catalytic rate, an altered ADPGPP substrate specificity, and an increased ability of a cell expressing one or more members of the library to produce starch when the one or more library members is expressed in the cell, thereby producing a selected library of recombinant first enzyme nucleic acid homologues; and,

40 30 (D) recursively repeating steps A-C one or more times, wherein the selected library of recombinant first enzyme nucleic acid homologues provides one or more of:

5

10

5

the one or more ADPGPP enzyme coding nucleic acid, the homologue thereof, or the one or more homologous first nucleic acid of step (A), wherein steps A-C are repeated until one or more members of the selected library produces an elevated starch level in a target recombinant cell when the one or more selected library member is expressed in the target cell, as compared to a natural starch level of the target cell when the one or more selected library member is not expressed in the target cell.

15

10

20. The method of claim 19, wherein the recombining step is performed in vitro, in silico or in vivo, or a combination thereof.

20

22. The selected library of claim 19.

25

15

23. The one or more selected library member of claim 19.

24. The diversified library of claim 19.

25. The target recombinant cell of claim 19.

26. A plant comprising the target recombinant cell of claim 25.

30

35

40

45

50

55

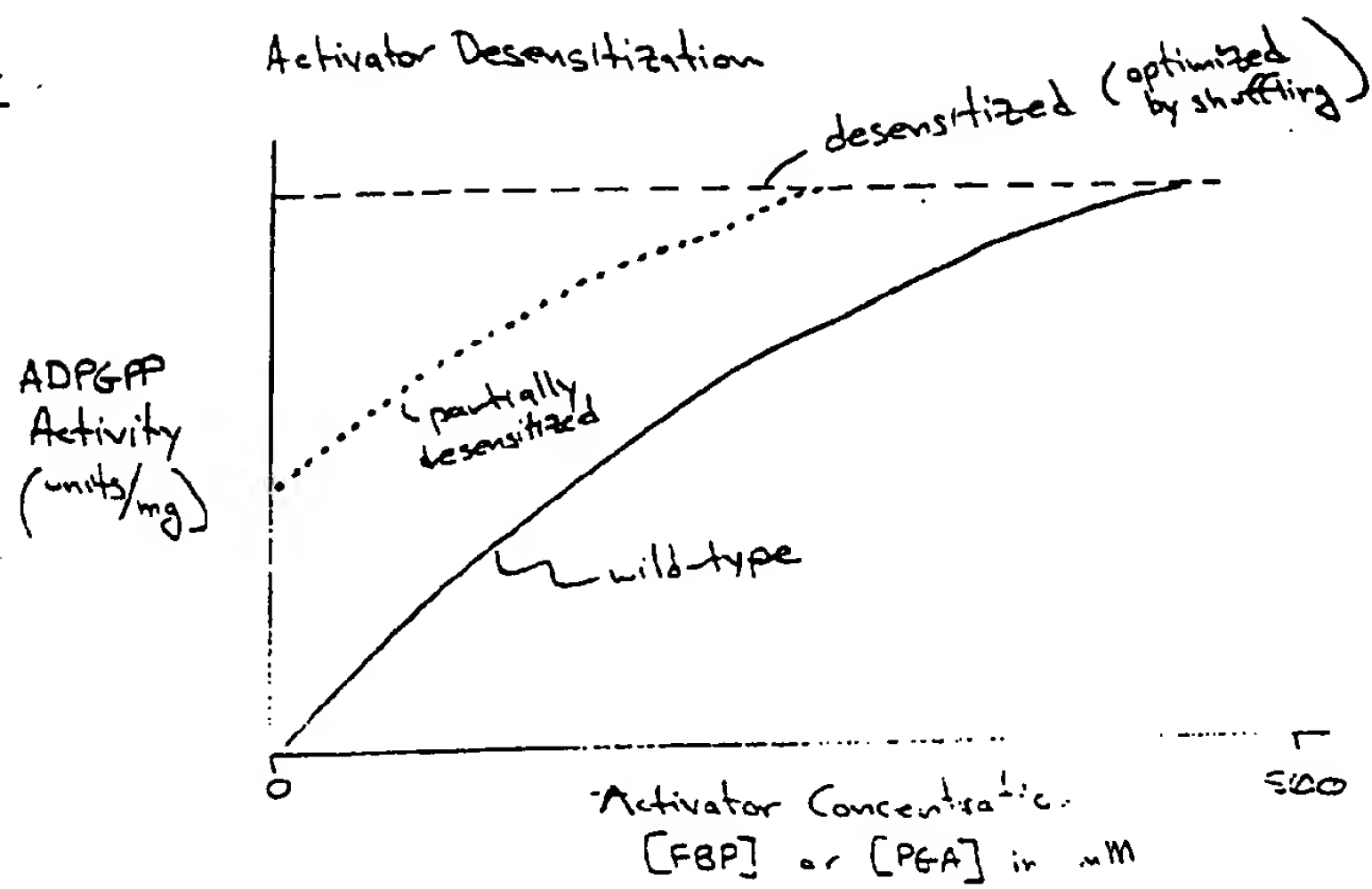
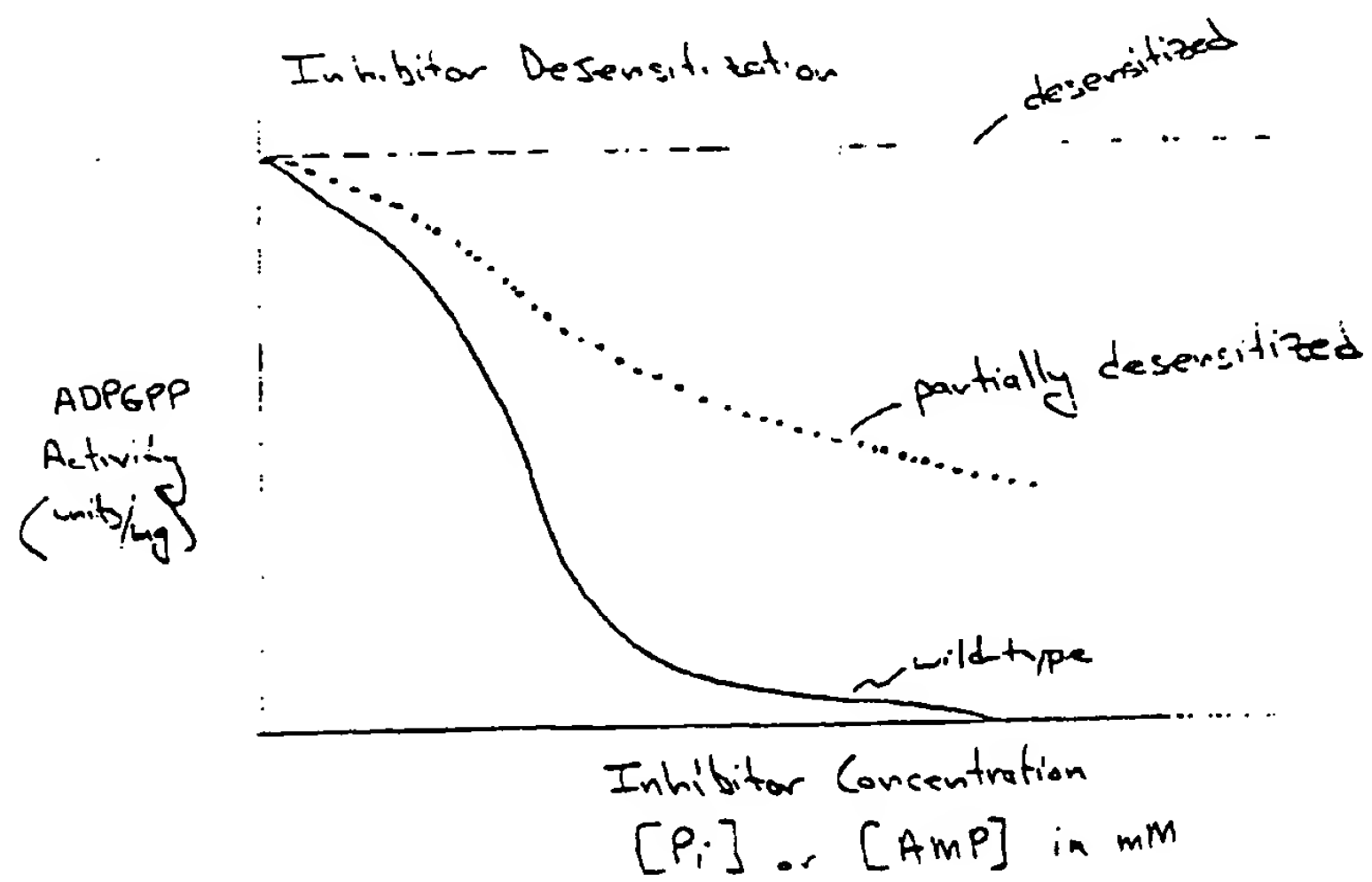
A.B.

FIGURE 1

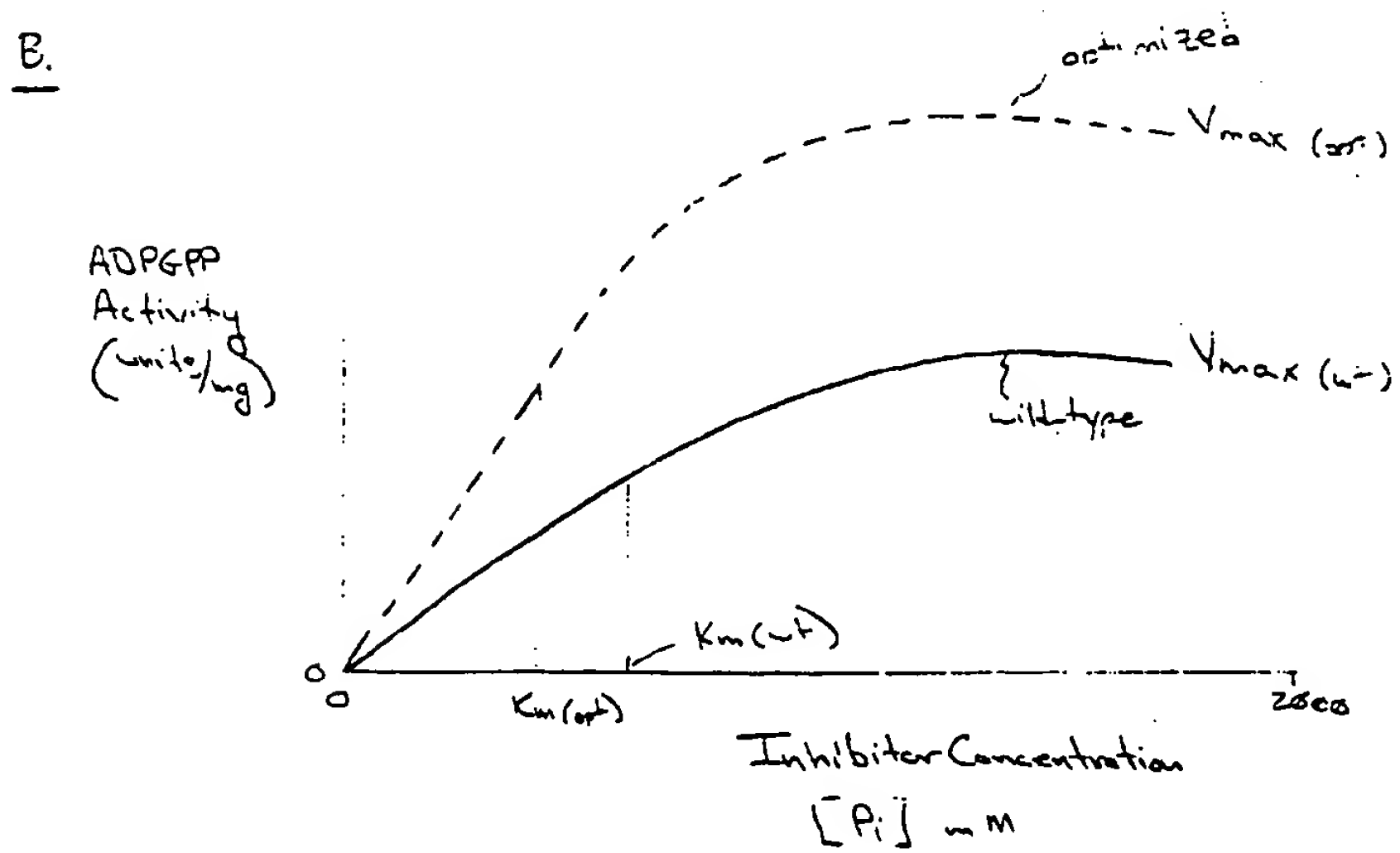
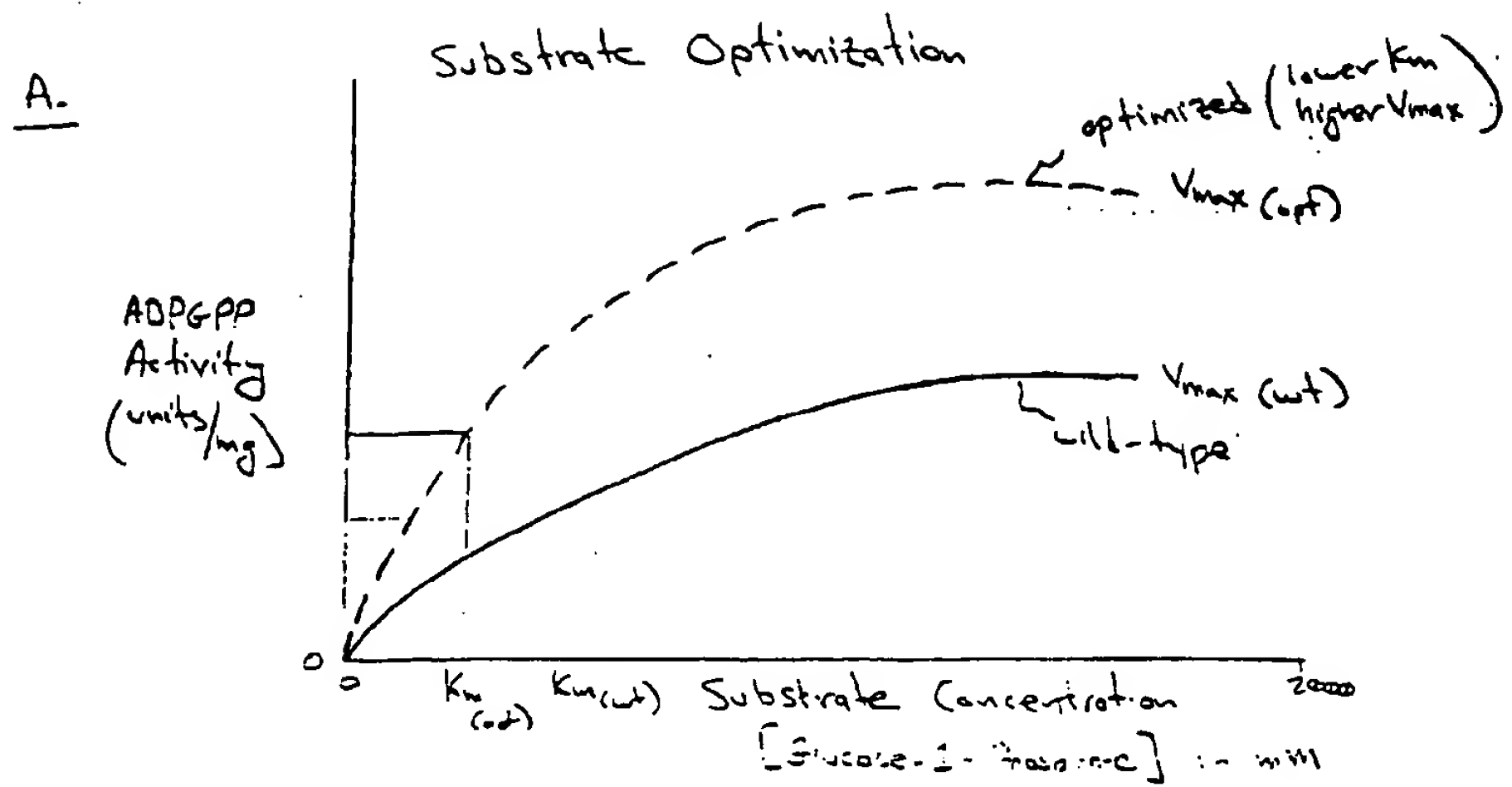


FIGURE 2

## INTERNATIONAL SEARCH REPORT

Int'l Application No  
PCT/US 99/26797

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/10 C12Q1/48 C12Q1/68 C12N5/10 A01H5/00  
C12N15/54

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CRAMERI A ET AL: "DNA SHUFFLING OF A FAMILY OF GENES FROM DIVERSE SPECIES ACCELERATES DIRECTED EVOLUTION" NATURE, GB, MACMILLAN JOURNALS LTD. LONDON, vol. 391, 15 January 1998 (1998-01-15), pages 288-291, XP000775869 ISSN: 0028-0836 the whole document — —/—	1-26

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the International filing date

"L" document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the International filing date but later than the priority date claimed

"T" later document published after the International filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"Z" document member of the same patent family

Date of the actual completion of the International search

21 March 2000

Date of mailing of the International search report

06/04/2000

Name and mailing address of the ISA

European Patent Office, P.B. 6818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Billang, J

## INTERNATIONAL SEARCH REPORT

 International Application No  
 PCT/US 99/26797

## C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	STEMMER W: "DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, US, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, vol. 91, 1 October 1994 (1994-10-01), pages 10747-10751, XP002087463 ISSN: 0027-8424 the whole document	1-26
Y	SMITH-WHITE B J ET AL: "COMPARISON OF PROTEINS OF ADP - GLUCOSE PYROPHOSPHORYLASE FROM DIVERSE SOURCES" JOURNAL OF MOLECULAR EVOLUTION, (MAY 1992) VOL. 34, NO. 5, PP. 449-464. ISSN: 0022-2844., XP000882154 the whole document	1-26
Y	T W GREENE ET AL: "Generation of up-regulated allosteric variants of potato ADP-glucose pyrophosphorylase by reversion genetics" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, US, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, vol. 95, no. 17, 1 August 1998 (1998-08-01), pages 10322-10327, XP002096426 ISSN: 0027-8424 the whole document	1-26
A	W0 94 24292 A (DANISCO ;VILLAND PER (NO); KLECZKOWSKI LESZEK (NO); OLSEN ODD ARNE) 27 October 1994 (1994-10-27) the whole document	
A	W0 98 41622 A (NOVONORDISK AS) 24 September 1998 (1998-09-24) the whole document	
A	HARAYAMA S: "Artificial evolution by DNA shuffling" TRENDS IN BIOTECHNOLOGY, GB, ELSEVIER PUBLICATIONS, CAMBRIDGE, vol. 16, no. 2, 1 February 1998 (1998-02-01), pages 76-82, XP004107046 ISSN: 0167-7799 the whole document	
	-/-	

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 99/26797

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>RUOJ H ET AL: "A (His)(6)-tagged recombinant barley (<i>Hordeum vulgare</i> L) endosperm ADP - glucose pyrophosphorylase expressed in the baculovirus-insect cell system is insensitive to allosteric regulation by 3-phosphoglycerate and inorganic phosphate"</p> <p>FEBS LETTERS, (8 DEC 1997) VOL. 419, NO. 1, PP. 124-130 ISSN: 0014-5793., XP000882835</p> <p>the whole document</p>	

# INTERNATIONAL SEARCH REPORT information on patent family members

Inter- national Application No  
PCT/US 99/26797

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9424292 A	27-10-1994	AU 693787 B	09-07-1998
		AU 6539294 A	08-11-1994
		CA 2160159 A	27-10-1994
		EP 0693128 A	24-01-1996
		GB 2291878 A, B	07-02-1996
		JP 8509121 T	01-10-1996
		NZ 265061 A	27-07-1997
		NZ 314961 A	25-02-1999
		US 5977437 A	02-11-1999
WO 9841622 A	24-09-1998	AU 6611498 A	12-10-1998